

Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2004, 09:34:43 ; Search time 142 Seconds
(without alignments)

2108 639 Million cell updates/sec

Title: US-09-671-687A-3

Perfect score: 5034

Sequence: 1 MSSGILWSQEKVTPSPYWEERI.....RLLCDAYCMYQSPTMSLYK 949

Scoring table: BL050622

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : QPTRMBL_25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rabbit:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archeap:*

Result No.	Score	Query	Match length	DB ID	Description
1	4983	99.0	960	4 O94334	O94334 homo sapien
2	98.8	██████████	956	4 Q9NQC7	Q9NQC7 homo sapien
3	4718.5	93.7	952	11 Q8CGCB0	Q8CGCB0 mus musculu
4	4718.5	93.7	999	11 Q80TQ2	Q80TQ2 mus musculu
5	2952	58.6	623	11 Q8BYL9	Q8BYL9 mus musculu
6	2512	49.9	515	4 Q9NQX9	Q9NQX9 homo sapien
7	1662	33.0	313	11 Q80VB3	Q80VB3 mus musculu
8	856	22.0	318	11 Q8IXX23	Q8IXX23 mus musculu
9	1136.5	22.6	551	5 Q8IPC3	Q8IPC3 drosophila
10	1136.5	22.6	639	5 Q8IPC5	Q8IPC5 drosophila
11	1134	22.5	550	5 Q8IPC4	Q8IPC4 drosophila
12	1132	22.5	639	5 Q8SF0	Q8SF0 drosophila
13	856	17.0	517	5 Q9VL04	Q9VL04 drosophila
14	548	10.9	970	5 Q9U3F9	Q9U3F9 caenorhabdi
15	149.5	3.0	1046	11 Q9z0H8	Q9z0H8 rattus norv
16	149	3.0	1046	11 Q9z0H8	Q9z0H8 mus musculu

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match length	DB ID	Description
1	4983	99.0	960	4 O94334	O94334 homo sapien
2	98.8	██████████	956	4 Q9NQC7	Q9NQC7 homo sapien
3	4718.5	93.7	952	11 Q8CGCB0	Q8CGCB0 mus musculu
4	4718.5	93.7	999	11 Q80TQ2	Q80TQ2 mus musculu
5	2952	58.6	623	11 Q8BYL9	Q8BYL9 mus musculu
6	2512	49.9	515	4 Q9NQX9	Q9NQX9 homo sapien
7	1662	33.0	313	11 Q80VB3	Q80VB3 mus musculu
8	856	22.0	318	11 Q8IXX23	Q8IXX23 mus musculu
9	1136.5	22.6	551	5 Q8IPC3	Q8IPC3 drosophila
10	1136.5	22.6	639	5 Q8IPC5	Q8IPC5 drosophila
11	1134	22.5	550	5 Q8IPC4	Q8IPC4 drosophila
12	1132	22.5	639	5 Q8SF0	Q8SF0 drosophila
13	856	17.0	517	5 Q9VL04	Q9VL04 drosophila
14	548	10.9	970	5 Q9U3F9	Q9U3F9 caenorhabdi
15	149.5	3.0	1046	11 Q9z0H8	Q9z0H8 rattus norv
16	149	3.0	1046	11 Q9z0H8	Q9z0H8 mus musculu

ALIGNMENTS

RESULT 1					
TID	094934	PRELIMINARY;	PRT;	960 AA.	
* AC	094934	Q96EB10;	DT	01-MAY-1999 (TREMBLrel. 10, Created)	
			DT	01-OCT-2001 (TREMBLrel. 18, Last sequence update)	
			DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)	
			DE	Hypothetical protein KIAA0849 (Cylinromatosis) (Turban tumor syndrome) (Fragment).	
			DE	KIAA0849	
			OS	Homo sapiens (Human).	
			OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Homo.	
			OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
			OX	NCBI_TaxID=9606;	
			RN	[1]	
			RP	SEQUENCE FROM N.A.	
			RC	TISSUE=brain;	
			RX	MEDLINE=9115230; PubMed=1048485;	
			RA	Nagase T., Ishikawa K., Suyama M., Kitkuno R., Hiroswa M., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.; "Prediction of the coding sequences of unidentified human genes."	
			PT	RT XII: The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro."	
			RT	RL DNA Res. 5:355-364(1998)	
			RT	[2]	
			RL	SEQUENCE OF 8-960 FROM N.A.	
			RN	Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.	
			DR	EMBL: AB020656; BRA74872.2.	
			DR	EMBL: BC012342; AAH12342.1.	
			DR	Genew; HGNC:2584; CYLD.	
			DR	GO: 0004157; F: cysteine-type endopeptidase activity; IEA.	
			DR	GO: GO:0006511; P: ubiquitin thioesterase activity; IEA.	
			DR	InterPro: IPR00938; CAP-Gly.	
			DR	InterPro: IPR01394; Peptidase_C19.	
			PFam;	PF01302; CAP_GLY_2.	
			PFam;	PF00443; UCH_1.	

DR	PROSITE; PSS0245; CAP_GLY_2; 2.
KW	Hypothetical protein.
FT	NON_TER 1
SQ	960 AA; 107911 MW; 6D38410E57A9B289 CRC64;
Query Match	99.0%; Score 4983; DB 4; Length 960;
Best Local Similarity	99.5%; Pred. No. 0;
Matches	948; Conservative 0; Mismatches 1; Indels 4; Gaps 4;
Qy	1 MSSGLNSQEKTSPYWEERIYFLLOQCSVTDKOTOKLKVPGSIGQYIOPRSVGHSR1 60 8 MSSGLNSQEKTSPYWEERIYFLLOQCSVTDKOTOKLKVPGSTGQYIOPRSVGHSR1 67
Db	61 PSAKGKRNQIGKILEQPHAVLFYDE-DVEINEKEFTELLAITNCEERPSLKFKRNRLS 119 68 PSAKGKRNQIGKILEQPHAVLFYDE-DVEINEKEFTELLAITNCEERPSLKFKRNRLS 127
Db	120 KQLQIDVGCPVKVQLRSGEKEPGVVRFRGPAELAERTVSGIFGVELLERGRGQGTGDV 179 128 KQLQIDVGCPVKVQLRSGEKEPGVVRFRGPAELAERTVSGIFGVELLERGRGQGTGDV 187
Db	180 YGKOLFOCDCG-FVALDKLEIEDDTDALESDYAGPQDMQVELPPLEINSRVSLKG 238
Qy	188 YGKOLFOCDCG-FVALDKLEIEDDTDALESDYAGPQDMQVELPPLEINSRVSLKG 247
Db	239 GETIESCTVIFCDVLPGKESLGKYLEGKMDNPIGNWDGRFDGQVLCFSACVETSLIHIN 297 248 GETIESCTVIFCDVLPGKESLGXFGVYDMDNPIGNWDGRFDGQVLCFSACVETSLIHIN 307
Qy	298 D1PEPSVTOERPPLKAFMSRGVGDKSSSHNPKATGSTSDPGNR-RSELEPYTNGSS 356
Db	308 D1PEPSVTOERPPLKAFMSRGVGDKSSSHNPKATGSTSDPGNRSELEPYTNGSS 367
Qy	357 DSQPOSKSNTWYIDEADPASKLBTISTDFRSSPPLQPPVNLTTENRFHSLPFL 416
Db	368 DSQPOSKSNTWYIDEADPASKLBTISTDFRSSPPLQPPVNLTTENRFHSLPFL 427
Qy	417 TPKMPTNIGSPLSVAQSMEELTAPVQESPLAMPENGNHGLEEVGSIAEVKENP 476
Db	428 TPKMPTNIGSPLSVAQSMEELTAPVQESPLAMPENGNHGLEEVGSIAEVKENP 487
Qy	477 FYGVIRNIGQPPCQNLNEVLAGLEDECACTGDTFRGTRYTCAKKALFKLKSRPDS 536
Db	488 FYGVIRNIGQPPCQNLNEVLAGLEDECACTGDTFRGTRYTCAKKALFKLKSRPDS 547
Qy	537 RFASLQPVSNQIERCNSLAFFGTYLSEYVENTPPKMEKEGLIMIGRKKGIGHYNSCYL 596
Db	548 RFASLQPVSNQIERCNSLAFFGTYLSEYVENTPPKMEKEGLIMIGRKKGIGHYNSCYL 607
Qy	597 DSTLFCCLFAESSLDTLIRPKENQDVYEYSETQELIRTEIWPRLYGYCATKMKLR 656
Db	608 DSTLFCCLFAESSLDTLIRPKENQDVYEYSETQELIRTEIWPRLYGYCATKMKLR 667
Qy	657 KILEKVEAASGFTSEEDPPEFNLILPHIURVEPLKIRISQKVHQDCYFQIMEKNE 716
Db	668 KILEKVEAASGFTSEEDPPEFNLILPHIURVEPLKIRISQKVHQDCYFQIMEKNE 727
Qy	717 KYGVPTIQLLMFSINSLKFAEAAPSCLLITOMPREFGKDFKLFLKKIPPSLEINITDLED 776
Db	728 KYGVPTIQLLMFSINSLKFAEAAPSCLLITOMPREFGKDFKLFLKKIPPSLEINITDLED 787
Qy	777 TPRQCRCIGGLAMYECRECYDDBDISACKIKOQCKTONTQVHLHPBLNKYKPNVSLPKD 836
Db	788 TPRQCRCIGGLAMYECRECYDDBDISACKIKOQCKTONTQVHLHPBLNKYKPNVSLPKD 847
Qy	837 LPDDWHRHGCIQPNMELFAVLCIETSHYAVFKYRGKDDSAWLPSMADRGONGFNI 896
Db	848 LPDDWHRHGCIQPNMELFAVLCIETSHYAVFKYGDSSAHLPSMADRGONGFNI 907
Qy	897 P0VTPCPGEYLKMSLEDLHSLSRRIQGARRLLCDAYMCYQSPTMISLYK 949
Db	908 P0VTPCPGEYLKMSLEDLHSLSRRIQGARRLLCDAYMCYQSPTMISLYK 960
Qy	414 FSLTKMPNTNGSIGHSPLSLAQSVMBELNTAPVQESPPLAMPNGSHGLEVGSIAEVKE 473

RESULT 2

Q9NQC7 PRELIMINARY; PRT; 956 AA.

ID Q9NQC7; PRELIMINARY; PRT; 956 AA.

AC Q9NQC7; PRELIMINARY; PRT; 956 AA.

DT 01-OCT-2000 (Tremblrel. 15, Created)

DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)

DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DB Cyd protein.

GN Homo sapiens (Human).

OC Eukaryota; Metzoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OC NCBI_TaxID:9606;

RN [1] RP SEQUENCE FROM N.A.

MEDLINE=202996617; PubMed=10835629;

RX RX BIGNEAU,G.; Brown C., BIGGS P.J., Lakhani S.R., Jones C., Hanssen J.,

RA Blair E., Siebert R., Turner G., Evans D.G.,

RA Schrandt C., Beaman F.A., Van Den Ouweland A., Halley D.,

RA Delpech B., Cleveland M.G., Leigh J., Leisti J., Rasmussen S.,

RA Wallace M.R., Penske C., Banerjee P., Olson N., Cheggar R., Merrett S.,

RA Leonard N., Huber M., Hohl D., Chapman P., Burn J., Swift S.,

RA Smith A., Ashworth A., Stratton M.R.;

RT Identification of the familial cylindromatosis tumor suppressor gene.",

RL Nat. Genet. 25:160-165 (2000). 

DR EMBL; AJ250014; CAB33533; 1;

CO GO:0005856; Cytoskeleton; NAS.

DR GO; GO:0004221; Ubiquitin thioesterase activity; NAS.

DR GO; GO:0006511; Ubiquitin-dependent protein catabolism; NAS.

DR InterPro; IPR00938; CAP-Gly.

DR InterPro; IPR01394; Peptidase_C19.

PFAM PF01302; CAP_Gly; 2.

DR PROSTB; PS50245; CAP_GLY; 2.

DR PROSITE; PS5025; UCH_2_3; 1.

SQ 956 AA; 107315 MW; 01811F9b83424631 CRC64;

Query Match 98.8%; Score 4971.5; DB 4; Length 956;

Best Local Similarity 99.2%; Pred. No. 0;

Matches 948; Conservative 0; Mismatches 1; Indels 7; Gaps 5;

Qy 1 MSSGLNSQEKTSPYWEERIYFLLOQCSVTDKOTOKLKVPGSIGQYIOPRSVGHSR1 60

Db 1 MSSGLNSQEKTSPYWEERIYFLLOQCSVTDKOTOKLKVPGSIGQYIOPRSVGHSR1 60

Qy 61 PSAKGKRNQIGKILEQPHAVLFYDE-DVEINEKEFTELLAITNCEERPSLKFKRNRLS 119

Db 61 PSAKGKRNQIGKILEQPHAVLFYDE-DVEINEKEFTELLAITNCEERPSLKFKRNRLS 120

Qy 180 YQKQLFQCDCEG-FVALDKLEIEDDTALESDYAGPGDTMQVQELPLEINSPVSLKG 238

Db 181 YQKQLFQCDCEGCFVYDQLKELIEDDTALESDYAGPGDTMQVQELPLEINSPVSLKG 240

Qy 120 KGLQIDVGCPVKVQLRSGEEKPGVYFRGPAELAERTVSGIFPGVYFRGCOGFTDGV 179

Db 121 KGLQIDVGCPVKVQLRSGEEKPGVYFRGPAELAERTVSGIFPGVYFRGCOGFTDGV 180

Qy 241 GETIESCTVIFCDVLPGKESLGXFGVYDMDNPIGNWDGRFDGTYQDGLCPVCFESTILLHIN 300

Db 298 DIPT--ESVTOERRPKLAFMSRGVYDGFQDCVTDPEKGRFDGTY-LCSPACVBSSTILLHIN 293

Qy 319 GETIESCTVIFCDVLPGKESLGXFGVYDMDNPIGNWDGRFDGTYQDGLCPVCFESTILLHIN 300

Db 301 DIPLALSESVTOERRPKLAFMSRGVYDGFQDCVTDPEKGRFDGTYQDGLCPVCFESTILLHIN 360

Qy 354 SVVDSDQSKSNTWYIDEVADPKSLTEISTDFBSSPPLQPPYVNSLTTENRFHSILP 413

Db 361 SSVDSDQSKSNTWYIDEVADPKSLTEISTDFBSSPPLQPPYVNSLTTENRFHSILP 420

GN MKIAA0849.	Db 587 RPASLQPVSNQIERCNSLAFFGGLSEVEENTPPKMEKEGLBIMIGKKKGIOGHYNCSYL 646
OS Mus musculus (Mouse).	Qy 597 DSTLFLCLFAFSVSLDVTLLRPEKEKNVEYSETQELRLTETVNPRLIYGYCATKIMKL R 656
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	Db 647 DSTLFLCLFAFSVSLDVTLLRPEKEKNVEYSETQELRLTETVNPRLIYGYCATKIMKL R 706
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX NCBI_TaxID=10090;	
RN [1]	
SEQUENCE FROM N.A.	
RC TISSUE=Brain;	
RX MEDLINE=227979291; PubMed=12693553;	
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,	
RA Nakajima D., Nagase T., Ohara O., Koga H.;	
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:	
II. The complete nucleotide sequences of 400 mouse KIAA-homologous	
CDNA's identified by screening of terminal sequences of cDNA clones	
RT randomly sampled from size-fractionated libraries.",	
RL DNA Res. 10:35-48 (2003).	
DR EMBL: AK122389; BAC65671_1; -.	
DR GO; GO:0004197; F:cytokeratine-type endopeptidase activity; IEA.	
DR GO; GO:0004221; F:ubiquitin thioesterase activity; IEA.	
DR GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.	
DR InterPro; IPR000938; CAP_Gly.	
DR InterPro; IPR001394; Peptidase_C19.	
DR InterPro; IPR000408; Reg_chr_Condens.	
DR Pfam; PF00443; CAP_GLY; -.	
DR PROSITE; PS50245; CAP_GLY_2; 2.	
DR PROSITE; PS50266; RCC1_2; 1.	
DR PROSITE; PS50235; UCH_2_3; 1.	
FT NON_TER 1 AA; 111979 MW; 8E2BF28CCAB22276 CRC64;	
SQ SEQUENCE 999 AA; 111979 MW; 8E2BF28CCAB22276 CRC64;	
RESULT 5	
Q8BYL9 ID Q8BYL9; PRELIMINARY; PRT; 623 AA.	
AC Q8BYL9; DT 01-MAR-2003 (T=EMBLrel. 23; Created)	
DR DT 01-MAR-2003 (T=EMBLrel. 23; Last sequence update)	
DR DT 01-OCT-2003 (T=EMBLrel. 25; Last annotation update)	
Qy 717 KILEVYEAASGETSEKDPEEFLNLDEHILRVEPLILKIRSAQKQVDCYFQIPEBKNE 716	
Db 707 KILEVYEAASGETSEKDPEEFLNLDEHILRVEPLILKIRSAQKQVDCYFQIPEBKNE 716	
Qy 757 KILEVYEAASGETSEKDPEEFLNLDEHILRVEPLILKIRSAQKQVDCYFQIPEBKNE 716	
Db 707 KILEVYEAASGETSEKDPEEFLNLDEHILRVEPLILKIRSAQKQVDCYFQIPEBKNE 716	
Qy 777 TPROCRICGGLAMYECRECYDDISAGKIKOFCRVCNTQVHLHPKBLNHCKNPVSLPKD 836	
Db 827 TPROQRICGGLAMYECRECYDDISAGKIKOFCRVCNTQVHLHPKBLNHCKNPVSLPKD 886	
Qy 837 LPDWDRHRGC1PCQNNMELFAVLCIESHYAFVYKQGDSSAWLFDFDSMADRGQGQNGENI 896	
Db 887 LPDWDRHRGC1PCQNNMELFAVLCIESHYAFVYKQGDSSAWLFDFDSMADRGQGQNGENI 946	
Qy 897 PQYTPCPBEVGTYLKMSELDLHSLSRRIQGARRLLCDAYCMYQSPTMISLYK 949	
Db 947 PQYTPCPBEVGTYLKMSELDLHSLSRRIQGARRLLCDAYCMYQSPTMISLYK 999	

Db	121	KGLQVDVGGSPVKVOLSGEEBKFPGVVRFFGPLAERTVSGIFFGVELLBEGRGQGFTDGV	180		61	GVRIGQPPLNEVLAGLEDECAAGCTDGTFRRTGYFTCAKALFVKLKSCRPDSSRF	120
Qy	180	YQKQLFQDDEDGCFVADLKKLEIEDDTALEDSDYAGGDDTMOVELPLEINSRVLKG	238		539	ASLQVPSNQLERCNLSAFCGSLFGYFLSEVEENTPPKMKMKEGLEIMIGKKGIGQHNSCYLDS	598
Db	181	YQKQLFQDDEDGCVFVADLKKLEIEDDTALEDSDFAGGDDTMOVEPPLIEINSRVLKG	240		121	ASLQVPSNQLERCNLSAFCGSLFGYFLSEVEENTPPKMKMKEGLEIMIGKKGIGQHNSCYLDS	180
Qy	239	GETI-LESGTIVFCDVLPKGESLGYFGVDMNDPIGNWDGRFDGV-LCSFACVESTILLHN	297		599	TLCFCFAFSSVLDTTLRPEKEKDVEYYSTQELARTEIVNPLRYGYVCATKIMKLRKI	658
Db	241	GESTESGTIVFCDVLPKGESLGYFGVDMNDPIGNWDGRFDGVLCFSAVESTILLHN	300		181	TLCFCFAFSSVLDTTLRPEKEKDVEYYSTQELARTEIVNPLRYGYVCATKIMKLRKI	240
Qy	298	DITP---ESEVTOQERRPKLAEMSRGFGDKGSSSHNPKPATGSTS DPGNR-RSELFYTLMG	353		659	LEKVEAAASGETSEEDPPEEFLNLPHHLRVEPLKIRSAQKVODCFCYFOIFMEKNEKV	718
Db	301	DITPALSDVTQERRPKLAEMSRGFGDKGSSSHNPKPATGSTS DPGNR-RSELFYTLMG	360		241	LEKVEAAASGETSEEDPPEEFLNLPHHLRVEPLKIRSAQKVODCFCYFOIFMEKNEKV	300
Qy	354	SSVSDSQPQKSNTWYIDEVAEPAKSLTEISTDFDRSSPLQOPPPVNSLTENRFHSIP	413		719	GVPTIQQLEWSFINSLNPKAEAPCLQTOMPRFKDKEFLFKKIPSLSLELNITDJEEDTP	778
Db	361	SSVDSQ-QSKSKNPWYIDEVAEPAKSLTEISMSSDFGHSSPPPQQPSVNSLSENRFHSIP	419		301	GVPTIQQLEWSFINSLNPKAEAPCLQTOMPRFKDKEFLFKKIPSLSLELYQIVYLKTLP	360
Qy	414	FSLTKMPNNTGSGHSPLSLSAQSVMEELTAPYQESPLAMPFGNSHCGLEYGSLAETKE	473		779	RQCRTIGGLAMYBCELCYDPPDISAGKIKOFCKTCNTQVHLHPKRLNHKNPNVSLPKDLP	838
Db	420	FSLTKMPNNTGSGHSPLSLSVQSMGEINSTPVQESPPLISSGNAHGLEVGSLAETKE	479		361	DSAGYVEGLQCMSSVENATTIRTSGAKIQFCKTCTNTQVHLHPKRLNHKNPNVSLPKDLP	420
Qy	474	NPPFYGVIRIWQGPPLNENVLAFLAGLEDECAAGCTDGTFFGTRYTCALKKALFVKKLKSQR	533		839	DWDWRHGCIPCQNMELFAVLICETSHYAVFKVQGDQDGSNLFDSMADRGQDQNGFNIPQ	898
Db	480	NPPFYGVIRIWQGPPLNENVLAFLAGLEDECAAGCTDGTFFGTRYTCALKKALFVKKLKSQR	539		421	DWDWRHGCIPCQNMELFAVLICETSHYAVFKVQGDQDGSNLFDSMADRGQDQNGFNIPQ	480
Qy	534	PDSRFASLCPVSNOIERCNSLAFGYSLSVVEENTPPKMEKEGLEIMIGKKGIGQHNS	593		899	VTPCPCEVGEYLKMSLDELIHSLSRSRI	924
Db	540	PDSRFASLCPVSNOIERCNSLAFGYSLSVVEENTPPKMEKEGLEIMIGKKGIGQHNS	599		481	VTPCPCEVGEYLKMSLDELIHSLSRSRI	506
Qy	594	CYLDSTLFCFLFCAFSVLDVLLRP	617				
Db	600	CYLDSTLFCFLFCAFSVLDVLLRP	623				
				RESULT 7			
				Q80VB3	SEQUENCE FROM N.A.		
				ID	Q80VB3;	PRELIMINARY;	
				AC	Q80VB3;	PRT;	313 AA.
				DT	01-JUN-2003	(TREMBLrel.	24, Created)
				DT	01-JUN-2003	(TREMBLrel.	24, Last sequence update)
				DT	01-JUN-2003	(TREMBLrel.	24, Last annotation update)
				DE		Hypothetical protein (Fragment).	
				OS		Mus musculus (Mouse).	
				OC		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
				OC		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
				OX		NCBI_TaxID=10090;	
				RN		[1]	
				RP		SEQUENCE FROM N.A.	
				RC		STRAIN=FVB/N; TISSUE=Breast tumor;	
				RA		Strausberg R.;	
				RL		Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.	
				EMBL;		Bc049879; AAH49879-1; -.	
				FT		Hydrothetical protein.	
				NON_TER			
				SEQUENCE	313 AA:	36198 MW;	BBD5662BAA7453DA CRC64;
				Query Match		33.0%;	Score 1662; DB 11;
				Best Local Similarity		97.8%;	Length 313;
				Matches	30;	Conservative 3;	Mismatches 4;
						Indels 0;	Gaps 0;
				Qy	637	IYNPLRYGYVCATKIMKLRKILLEKEAASGFTSEKDPBEFINTLFFHILRVEPLLKIR	696
				Db	1	IYNPLRYGYVCATKIMKLRKILLEKEAASGFTSEKDPBEFINTLFFHILRVEPLLKIR	60
				697	SAGOKYODCFCYFOIFMEKNEVYGVYPYTIQQLLENFINSNLKAPAPSCLIJOMPRFGKDF	756	
				Qy	61	SAGOKYQDNFNYQQLFMEKEKVGYPTIQLLENFINSNLKAPAPSCLIJOMPRFGKDF	756
				Db	121	KLPKKIPSLSLAERTVSGIFFGVELLBEGRGQGFTDGV	180
				817	VHLHPKRLNHKNPVSPLPKDLPDWHRHGCIPCMELFAVLCLTSHTYAFVKYGDSS	876	
				Db	181	VHLHPRRLNHSYHFVSLPKDLPDWHRHGCIPCMELFAVLCLTSHTYAFVKYGDSS	240
				Qy	479	GVIRIGQPPLNEVLAGLEDECAAGCTDGTFRRTGYFTCAKALFVKLKSCRPDSSRF	538

Qy	877	AWLFDSDMADRGGGNGENIPQVTFPCPEGYLMNSLEDLHSLSRRIQGCARRILCDAY	936
Db	241	AWLFDSDMADRGGGNGENIPQVTFPCPEGYLMNSLEDLHSLSRRIQGCARRILCDAY	300
Qy	937	MCMYQSPTMNSLYK	949
Db	301	MCMYQSPTMNSLYK	313
RESULT 8			
	QB8XZ3	PRELIMINARY;	PRT;
	AC	318 AA.	
	DT	01-MAR-2003 (TREMBrel. 23, Created)	
	DT	01-MAR-2003 (TREMBrel. 23, Last sequence update)	
	DT	01-JUN-2003 (TREMBrel. 24, Last annotation update)	
	DE	CYLINDROMATOSIS.	
	GN	C13039D01RK.	
	OS	Mus musculus (Mouse).	
	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
	NCBI_TaxID	=10090; [1]	
	RN	RP	SEQUENCE FROM N.A.
	RC	STRAIN=C57BL/6J; TISSUE=Cerebellum;	
	RX	MEDLINE=22354683; PubMedId=12466851;	
	RA	The FANTOM Consortium,	
	RA	the RIKEN Genome Exploration Research Group Phase I & II Team;	
	RT	"Analysis of the mouse transcriptome based on functional annotation of	
	RT	60,770 full-length cDNAs.";	
	RL	Nature 420:563-573 (2002).	
	DR	EMBL; AK042764; BAC31357.1; -	
	DR	MGD; MGJ:2445104; C13009D01Rik.	
	DR	InterPro: IP0000938; CAB-Gly.	
	DR	PFAM: PF01302; CAP_GLY_1.	
	DR	PROSITE: PS50454; CAP_GLY_2; 1.	
	DR	SEQUENCE: 318 AA; 35258 MW; 23465D36304356BF CRC64;	
	Query Match	29 /8;	Score 1458.5;
	Best Local Similarity	94.1%;	DB 11;
	Matches	286; Conservative 5; Mismatches 10; Indels 3;	Length 318;
		Gaps 3	
Qy	1	MSSGMWSQEKVTPSPWEEFYLQIQLQCSVTDKOTQKLKVPKGSGTQYDTSYQGHSR1	60
Db	1	MSSGMWSQEKVTPSPWEEFYLQIQLQCSVTDKQKLKVPKGSGTQYDTSYQGHSRV	60
Qy	61	PSAKGKGKQIGLKILQPHAVIYDFE-DVYVINEKFTELLIAITNECEERFSLEKPNRPLS	119
Db	61	PSTRCKKGKQIGLKILQPHAVIYDFE-DVYVINEKFTELLIAITNECEERFSLEKPNRPLS	120
Qy	120	KGLQDVGCPVVKVQRSGEKXPGVFRGPILLAERTVSGLFFGGELLEERGGFTDGV	179
Db	121	KGLQDVGSPVKVQRSGEKXPGVFRGPILLAERTVSGLFFGGELLEERGGFTDGV	180
Qy	180	YQGRKLFQCDBEDCG-FVALDKLIELBDDDTALESDYAGPQDPMQYELPPLBINSRVSLKG	238
Db	181	YQGRKLFQCDBEDCGFVALDKLIELBDDDGLESDFAGPQDMQYELPPLBINSRVSLKG	240
Qy	239	GETIHSQTVIICDVLPKGKESLIGYFYGVDMNPIGNWDGRFDGV-LCSFACVESTLILHIN	297
Db	241	GETIHSQTVIICDVLPKGKESLIGYFYGVDMNPIGNWDGRFDGVQLCFASTESTLILHIN	300
Qy	298	DIIP 301	
Db	301	DIIP 304	
RESULT 9			
	QB1PC3	PRELIMINARY;	PRT;
	ID	551 AA.	
	OB1PC3		
	AC		
	DT	01-MAR-2003 (TREMBrel. 23, Created)	

SEQUENCE FROM N.A.	
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,	Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,	Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RA "Annotation of Drosophila melanogaster genome"; [4]	Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RL SEQUENCE FROM N.A.	SEQUENCE FROM N.A.
RA Adams M.D., Celikiner S.E., Gibbs R.A., Rubin G.M., Venter C.J.,	Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RA [5]	SEQUENCE FROM N.A.
RP Flybase;	SEQUENCE FROM N.A.
RA Submitted; (SEP-2002) to the EMBL/GenBank/DDBJ databases.	Submitted; (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RA EMBL; AE03628; AAN0738..1..;	EMBL; AE03628; AAN0738..1..;
RA Flybase; FBgn0032210; CG5603.	Flybase; FBgn0032210; CG5603.
DR GO; GO:0004197; F: cysteine-type endopeptidase activity; IEA.	DR GO; GO:0004197; F: cysteine-type endopeptidase activity; IEA.
DR GO; GO:0004221; F: ubiquitin thioesterase activity; IEA.	DR GO; GO:0006511; P: ubiquitin-dependent protein catabolism; IEA.
DR InterPro; IPR001394; Peptidase_C19.	DR InterPro; IPR001394; Peptidase_C19.
PFam; PF00443; UCH; 1.	PFam; PF00443; UCH; 1.
SQ SEQUENCE 639 AA; 72128 MW: 6E29D3B09FF5E55B CRC64;	SEQUENCE 639 AA; 72128 MW: 6E29D3B09FF5E55B CRC64;
Query Match 22.6%; Score 1136.5; DB 5; Length 639;	Query Match 22.6%; Score 1136.5; DB 5; Length 639;
Best Local Similarity 41.3%; Matches 255; Conservative 103; Mismatches 202; Indels 57; Gaps 17;	Best Local Similarity 41.3%; Matches 255; Conservative 103; Mismatches 202; Indels 57; Gaps 17;
Qy 352 NGSSVDSPQSKSKNTWYIDEVAEDPAKSLLTEI-STDFDRSSPLQQPPVNLTTENRFH 410	Qy 352 NGSSVDSPQSKSKNTWYIDEVAEDPAKSLLTEI-STDFDRSSPLQQPPVNLTTENRFH 410
59 NHSNDVNQ-----HLDEV-----DLADTLKMPWPKAAMLNNSKTD--- 99	59 NHSNDVNQ-----HLDEV-----DLADTLKMPWPKAAMLNNSKTD--- 99
Db 411 SLPPFSLTMRPNTNGSIGHSPPLSLSAQSMVEINTAPQESPLAMPGNSHLEVGSLAE 470	Db 411 SLPPFSLTMRPNTNGSIGHSPPLSLSAQSMVEINTAPQESPLAMPGNSHLEVGSLAE 470
100 -----PSNSVDSLILKPASPLKIEPEEPRLRFTADYQPLIEITGE-LAIGSLVE 148	100 -----PSNSVDSLILKPASPLKIEPEEPRLRFTADYQPLIEITGE-LAIGSLVE 148
Qy 471 YTKNP---PFGVIRWMGPPG-LNEVLGLAHGLEDEC---AGCTPDGTRGTRYFTCAL 521	Qy 471 YTKNP---PFGVIRWMGPPG-LNEVLGLAHGLEDEC---AGCTPDGTRGTRYFTCAL 521
Db 149 V-SNPVCEDLGVRVMGPPQRNVLVGEVEDESNLKVVASGRHNGYRLFICHD 207	Db 149 V-SNPVCEDLGVRVMGPPQRNVLVGEVEDESNLKVVASGRHNGYRLFICHD 207
Qy 522 KKALFYKLKSRPDSRFASLQP--VSQNIERCNLSLAGGYLSEVVEVENTPKM--EKEGL 577	Qy 522 KKALFYKLKSRPDSRFASLQP--VSQNIERCNLSLAGGYLSEVVEVENTPKM--EKEGL 577
Db 208 GRAIFVBAANRCTADRRADPVISNANRVSNSNHAKKPGVADCPAIYGSIPPLQIHNNSDEL 267	Db 208 GRAIFVBAANRCTADRRADPVISNANRVSNSNHAKKPGVADCPAIYGSIPPLQIHNNSDEL 267
Qy 578 EIMGRKGKGQHYNNSCYLDLFCDAFSSLVYDPLVLRPREKNDVYESYSETQELRTEI 637	Qy 578 EIMGRKGKGQHYNNSCYLDLFCDAFSSLVYDPLVLRPREKNDVYESYSETQELRTEI 637
Db 268 ASICGFKKGQHYNNSCYLDLFCDAFSSLVYDPLVLRPREKNDVYESYSETQELRTEI 327	Db 268 ASICGFKKGQHYNNSCYLDLFCDAFSSLVYDPLVLRPREKNDVYESYSETQELRTEI 327
Qy 638 VNPLRIGYVCATKIMKLKLEKVEAASGFTSEEKDPEEFUNLPHHLRVPLKIRS 697	Qy 638 VNPLRIGYVCATKIMKLKLEKVEAASGFTSEEKDPEEFUNLPHHLRVPLKIRS 697
Db 328 VPLRKQNFVFRDRVMGLREIJDQLSSVSGLTCERDPEEFINSLSQIMRVEFLKSS 387	Db 328 VPLRKQNFVFRDRVMGLREIJDQLSSVSGLTCERDPEEFINSLSQIMRVEFLKSS 387
Qy 698 AGQKQDCCYFQIQLMEKNEKGYPTIQOLLEMSFINSNLKFAEAPSCLLIQMREFGKDFK 757	Qy 698 AGQKQDCCYFQIQLMEKNEKGYPTIQOLLEMSFINSNLKFAEAPSCLLIQMREFGKDFK 757
Db 368 G----QDSYFQLFVVERDEKLTPSVQLOFEQSFHSSDQKIKLKEVPSCTIIQMREFGKDVYK 443	Db 368 G----QDSYFQLFVVERDEKLTPSVQLOFEQSFHSSDQKIKLKEVPSCTIIQMREFGKDVYK 443
Qy 758 IFFKKFISLLENITDLDTPHQCRICTECERYDDPDISAG-KIKOFCKTNTQ 816	Qy 758 IFFKKFISLLENITDLDTPHQCRICTECERYDDPDISAG-KIKOFCKTNTQ 816
Db 444 MYPRILPSQVLDVTIDENSPOCSLCKLAAECRDFGSGLECTAFCPKCLKT 503	Db 444 MYPRILPSQVLDVTIDENSPOCSLCKLAAECRDFGSGLECTAFCPKCLKT 503
Qy 817 VHLPKPKRLNKHYNPVSLPKDLDWDWRHGCIPCONMELFATVLCIETSHYAFVWKYGR-DD 875	Qy 817 VHLPKPKRLNKHYNPVSLPKDLDWDWRHGCIPCONMELFATVLCIETSHYAFVWKYGR-DD 875
Db 504 FHSHIKEETNHVSKKYSPEKFIMA-BHMVVRLYMELFAVYCIETSHYAFVWKYGR-DD 562	Db 504 FHSHIKEETNHVSKKYSPEKFIMA-BHMVVRLYMELFAVYCIETSHYAFVWKYGR-DD 562
Qy 876 SAWLFFPSMADRGQGONGNFIPOVTPCPEVEYLK---MSLEDLHSLSRRIQGARRL 931	Qy 876 SAWLFFPSMADRGQGONGNFIPOVTPCPEVEYLK---MSLEDLHSLSRRIQGARRL 931
Db 563 APWCFPSMADRGKGEQNYNIPITCYVPLHEHAKRI 621	Db 563 APWCFPSMADRGKGEQNYNIPITCYVPLHEHAKRI 621
Qy 932 LCDAYMCMYQSPIMSLY 948	Qy 932 LCDAYMCMYQSPIMSLY 948
Db 622 FCDAYMLYQSTDIMMY 638	Db 622 FCDAYMLYQSTDIMMY 638
RN	RN
RP SEQUENCE FROM N.A.	RP SEQUENCE FROM N.A.
RA Celikiner S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y., Banzon J., Baldwin D., Bannister J., Besson J., Busam D.A., Carlson J.W., Center A., Champé M., Davenport L.B., Dietz S.M., Dodson K., Dorsett V., Douc L.E., Doyle C., Dresnek D., Farfan D., Ferreira B., Gallo R.F., Garg N.S., George R.A., Gonzalez M., Houck J., Hoskins R.A., Howland T.J., Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A., McIntosh T.C., Moy M., Murphy L., Muzy N.M., Nelson D.L., Nelson D.R., Nelson G.S., Pan S., Pollard J., Puri V., Reese M.G., Reiner K., Remington K., Saunders R.D., Scheier P., Shem H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svartkas R., Turner C., Turner E., Venter C., Weissenbach J., Wang Z.Y., Wasserman G.M., Weissenbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.F., Zaveri J.S., Zhang M., Zhao Q., Zheng L., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C., Williams S.M., Zaveri J.S., Smith H.O., Venter C., Tyler D., Rubin G.M., Stipelman M., Stroud R., Svarts R., Tector C., Smith H.O., Venter J.C., Rubin G.M., Scheeler F., Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F., Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M., "Sequencing of Drosophila melanogaster genome.", Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RP SEQUENCE FROM N.A.
RA Misra S., Crobb M.A., Matthews B.B., Bayraktaroglu L., Campbell K., Hrdlicky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D., Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M., Stipelman M., Stroud R., Svarts R., Tector C., Smith H.O., Venter J.C., Rubin G.M., "Sequencing of Drosophila melanogaster genome.", Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RP SEQUENCE FROM N.A.
RA [3]	RN

RA	Tupy J.L., Bergman C., Berman B., Carlson J.W., Ceuliniker S.E., Clamp M., Drysdale R., Emmert D., Fries E., de Grey A., Harris N., Kronmiller B., Marshall B., Millburn G., Richter J., Russo S., Searle S.M.J., Smith E., Shu S., Smutnicki F., Whitfield E., Ashburner M., Geibart W.M., Rubin G.M., Mungall C.J., Lewis S.E., RT	"Annotation of <i>Drosophila melanogaster</i> genome,"; Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[4]		
RP	SEQUENCE FROM N.A.	
RA	Adams M.D., Ceuliniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.; Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	
RN	[5]	
SEQUENCE FROM N.A.		
FLYBase;		
RA	Submitted (SEB-2002) to the EMBL/GenBank/DBJ databases.	
RL	EMBL; AE00328; AA10740; 1.	
DR	FlyBase; FBgn0032210; CG5603.	
GO	GO:0004197; F: cysteine-type endopeptidase activity; IEA.	
DR	GO; GO:0006521; F: ubiquitin thioesterase activity; IEA.	
DR	InterPro; IPR013394; Peptidase_C19.	
DR	Pfam; PF00443; UCH; 1.	
SQ	SEQUENCE 550 AA; 62083 MW; C87C73651D88FD59 CRC64;	
Query Match	22.5%; Score 1134; DB 5; Length 550;	
Best Local Similarity	45.7%; Pred. No. 1.5e-80;	
Matches	238; Conservative 87; Mismatches 167; Indels 28; Gaps 13;	
Qy	447 VQESPLAMPGNSHGLEVGSIAEVKENP---PFYGYTRWIGQOPPG-LNEVLAGLEED 501	
Db	38 IADYQPLIEIPGTB--LAIGSLIYEV-SNPQGVCDLYGTRWICTPPPGPOQNVLVGIEVED 94	
Qy	502 EC---AGCTDGTERGTRYFTCAKKAFLVKLKSRCRPSRFASLQP--VSQNQIERCRNSLA 555	
Db	95 ESMLNKVVAASDGRHNGVLFRTCHDGRAFVPARCTADRFRADVDNSISANRVSSSNHAKK 154	
Qy	556 FGGYLSEVVVEENTPPKM--EKEGLIMIGKKKGIGQHYNSCYLDSTLFCLEAFFSSVLDT 613	
Db	155 FGYADCPAIGSPTPQLOIHSDELASICKFKGQHNSCYLDATLMSMFTTSVPESI 214	
Qy	614 LURPKKEKNDVEYSETQELRLTTEIVNPRLIYGVCATKIMURKILKELEKVEAASGFTSREK 673	
Db	215 LYRRPGPQDINRYSEQVKVLRDEIVNPLRKVNVRSDRMVKLRELLDOLSSVSGLTCEEK 274	
Db	674 DPBEFLNTLFHFLTRVEPLKLIRSAGQKVQDCYFYQIFMFKNEKVGVPTIOOLLEWSFIN 733	
Db	275 DPEEFLSLSLQIMRVEPLKUSS---QDSYFQLEVKEDKTLPSVQQLFEQSFSHS 330	
Qy	734 SNIKEAFAPSCSLIIQMPRFKGDFKLFKKIPEPSLELNITDLDTPRQRICGGlamYECR 793	
Db	331 SDIKLKKEVPSCFIOMPFGKMYKMPFLPSQYDVTIDTIENSPROQSLCGKLAEBECR 390	
Qy	794 ECYDDPDISAG-KIKQFCKTCTNQVHLHPKLNHKYNPVSLLPKDLPDWWRHGCIPCONM 852	
Db	391 DCFGLSLQAGSGLECTAFCPKCLKTPHSHIKRTHVSKKYSKEPEKIMA-EHMVVPRLYM 449	
Qy	911 --MSLEDLHSLSRSRIOQCARILLCDAYCMYQSPIMSLY 948	
Db	510 EGARSINET-STDNDKVLPEHAKPIFCDAYMLYQSTDIMY 549	
Qy	853 ELFAVLCTESVHYAFVYKGK-DDSANLFDMSADRGQONGENIPQVTPCPEVGEYLK- 910	
Db	450 ELFAVVCLTESVHYAFVYKGSGDAPWCFDSNAKPCFKLTKPHSHIKRTHVSKKYSKEPEKIMA-EHMVVPRLYM 509	
Qy	504 FHSHIKRTHVSKKYSKEPEKIMA-EMVNPVLYMELFAVVCJETSHTYAFVKG-SGPD 562	
RESULT 12		
Q8SYFO	PRELIMINARY;	
ID	O8SYFO;	
AC	Q8SYFO;	
DT	01-JUN-2002 (TREMBLrel. 21, Created)	
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)	
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)	
DE	RE64280p.	

GN	CG5603.	
OS	<i>Drosophila melanogaster</i> (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;	
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC	Ephidoidea; Drosophilidae; Drosophila.	
OX	NCBItaxID:7227;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=Berkeley;	
RA	Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J., Champ M., Chavez C., Dorsett V., Dresnick D., Farfan D., Frise E., George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G., Mirande A., Mungall C.J., Nunoo J., Paragas V., Park S., Patel S., Phouanavong S., Wan K., Lewis S.B., Rubin G.M., Ceuliniker S.;	
RL	Submitted (DBE-2001) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AE00328; AA10740; 1; AA149214; 1.	
DR	Flybase; FBgn0032210; CG5603.	
GO	GO:0004197; F: cysteine-type endopeptidase activity; IEA.	
DR	GO; GO:0006521; F: ubiquitin thioesterase activity; IEA.	
DR	GO:0006511; F: ubiquitin-dependent protein catabolism; IEA.	
DR	InterPro; IP001334; Peptidase_C19.	
DR	Pfam; PF00443; UCH; 1.	
SQ	SEQUENCE 639 AA; 6216 MW; 65229D3BB757575B CRC64;	
Query Match	22.5%; Score 1132.5; DB 5; Length 639;	
Best Local Similarity	41.2%; Pred. No. 2.6e-80;	
Matches	254; Mis matches 203; Indels 57; Gaps 17;	
Qy	352 NGSSYDQFQSKSKNTWYIDEVAEDPAKSLTEI-STDFDRSSPPLQQPPVNSLTTENRFRH 410	
Db	59 NHSADVNZ-----HLDVD-----DLADLQTNPKRGAPEAMILANKSKTD---- 99	
Qy	411 SLPSLTKNPNTNGSIGHSPSLSAQSWEELNTAPQVESPPLAMPQGNSHGLEVSLAE 470	
Db	100 -----PSNSYDILKPKASPILKIEBPLRFTIADYQPLIBPGTB--LAIGSLV 148	
Qy	471 VKENP---PFGYVTRWIGQOPPG-LNEVLAGLEELDEC---AGCTDGTERGTRYFTCAL 521	
Db	149 V-SNPGVCDLYGTRWIGQPPQKVQNVLYGIEVEDESNLKVNVVASDGRHNGVRIFTCHD 207	
Qy	522 KKALFVLUKLSKCRPSRFASLQP--VSQNLERCRNSLAFGGYLSEVYVENTPPKM--EKEGL 577	
Db	208 GRAFIVPANRCTABRFRDVDNSTSANRVSNNHAKFGVADCPAIGSIPPLQHNSDEL 267	
Qy	578 EIMIGKKGIGOHYNNSCYLDSTLFCLEAFFSSVLDT 637	
Db	268 ASTIGKFKGIGHNNSCYLDATLMSMFTTSVDFSLVYRPGPOIRNTSEVQKVLRDEI 327	
Qy	638 VNPLRIYGVYCATKIMURKILKELEKVEAASGFTSREK 697	
Db	328 VNPFLKRYVFRSDRMVKLRELLDOLSSVGLTCEKDPEEFLNSLQIMRVEPLKLSS 387	
Qy	698 AGQKVKQDCYFYQIFMFKNEKVGVPTIQOLLEWSFIN 757	
Db	444 MYPFLIPSLQYDVTIDTIENSPROQSLCGKLAETVCRDCFGSLQAGSGLECTAFCPKCLKT 503	
Qy	817 VHJHPKRNHKNPVSLLPKDLPDWWRHGCIPCONMELFAVLCTESHTYAFVKG-DD 875	
Db	504 SANLFDSMADRGQNGFNIPQVTPCPEVGEYLK---MSLEDLHSLSRSRIOQCCARRL 931	
Qy	876 SANLFDSMADRGQNGFNIPQVTPCPEVGEYLK---MSLEDLHSLSRSRIOQCCARRL 931	
Db	563 APNCFDSMADRKGEQNGNIPETCVPRLTONSEEGRSINTI-STNDKVLPHAKR 621	
Qy	932 LCDAYMCYQSPTMSLY 948	
Db	622 FCDAYMCYQSTDIMY 638	

Qy	208 TALESDYACPGDTMQLVPPLEI-----NSRVS-LKGG-----ET 241	OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
Db	129 IREFSNYGRREEPCVIEIPGTMVYREMADDWKMSLELKEWFTEKRASSHLRDGLAMPME 188	NCBI_TaxID=10116;		
Db	129 IREFSNYGRREEPCVIEIPGTMVYREMADDWKMSLELKEWFTEKRASSHLRDGLAMPME 188	SEQUENCE FROM N.A.		
Qy	242 IESGTVIPE-----DVLD-GKEELGYFPGDMNPIGNWDGFGLVLSFACEVSTIL 294	LINE=9807115; PubMed=9427243;		
Db	189 LDC-TPLICAMITRSDVNRINODQIAHLLAVSVEKRI-----EVTNFENWNFLN 239	RX de Zeeuw C.I.; Hoogenraad C.C.; Goedknegt E.; Hertzberg E., RA Neubauer A.; Grossveld F.G.; Galjart N.J.; RT "CLP-115", a novel brain specific cytoplasmic linker protein, mediates the localisation of differential lamellar bodies.;"		
Qy	295 HNDIPEEV- TOERRPKPLNSRGYCDKG-----SSSHNKP 331	RT RL Neuron 19:1187-1198(1997).		
Db	240 KIGDVSVEVDETMRRVPPKWSWGERPPASGIWYNVDGDNTSQWPSSNQSYSSHDRL 299	DR EMBL; AJ000485; CAR041221; -.		
Qy	332 -----KATGSTS-DPGNRSELSEFTLN-----GSSVDS-----358	DR PIR; T42734; T42734.		
Db	300 NRQEDTNNWFEMCGSSSSVTPSN--SRLYSPNQNMMPMRGGGTSALYDNRRLQYSGDE 357	DR GO; GO:0005634; C: nucleus; IEA.		
Qy	359 -----QPSQSKSNTWYIDEVAEDPAKSLEISTDFDR-SSPPLQP 397	DR GO; GO:0003705; C: transcription factor complex; IEA.		
Db	358 QYRSAPKAPPERRIIPVSRQQPETEQRN-----SRSMKPSEPPDNTYSTHPPRP 406	DR GO:0006355; P: regulation of transcription, DNA-dependent; IEA.		
Qy	398 PPVNLST--TENEFHSLPSLTKPNT--NGSIGHSPPLSASQSYVMEELNTAPQESP 452	DR InterPro; IPR000938; CAP-Gly.		
Db	407 PPSSSMNVPYFSMSNTHSLOPSRSKSVQTIQRNQFTRQNPARNERVESQLNFRIGDQC- 464	DR InterPro; IPR000812; TFIIB_euk.		
Qy	453 LAMPGPNSHGLEVGSLAEVKENPFPYGVWIGOPPGLINEVLAGLEDECAGCTDGTRP 512	DR Pfam; PF1102; CAP_Gly; 2.		
Db	465 -----IWNNGAAE-----RGITKIVGFLKGHTKLTYAGVEFKN-TICAGTGYN 507	DR PROSITE; PS50245; CAP_Gly; 2.		
Qy	513 GTRYFTCAKKALFVKLKSCR-PDSRFASLQPVSNOIERTCNSLAFCGGTSEVYVENTTPK 571	DR PROSITE; PS00072; TFIIB_1.		
Db	508 REQLFLAKDGHAFVELSSLESFESSSISSTSSSSQHRR-----RLOSSRSQQMB- 557	DR SEQUENCE 1046 AA; 115480 MW; 72E669F76D2A1D0 CRC64;		
Qy	572 MEKGGLELMIGKKGKGIQHYNSCYLDSTFLCL-----FAFSVSLDVTLLRKEKNYD 624	Query Match 3.0%; Score 149.5; DB 11; Length 1046;		
Db	558 -AASGTS-SV---PVNGRHNVNGKQAS-ESLSDPPPPYYAPSPPRASLERSIKGS-E 611	Best Local Similarity 20.9%; Pred. No. 0.021;		
Qy	625 YSETQELLRTIEVNPLRIGYCATKIMKLRLK-L-EKVEAAGFTSEKDPPREFLNL 683	Matches 115; Conservative 66; Mismatches 189; Indels 179; Gaps 24;		
Db	612 TAQOFQKTLAHELVPLRKVHYRADHMKLRLKLAELMPHVTGLTNEEKDPEELTGF 671	Qy 309 RPPKLAFLMSRQGVGDGKSSSHNKPATGSTS DPGNRSELSEFTLNSSVDSQ-----PQ 361	DRSSSPPLQPPVNSLTTENR----409	
Qy	684 HHLRVPEPLKIRSAGQKVQDCFYQIPIFMEKNEVKGVTIQQLEWSFTNSNLKFAEAAPS 743	Db 8 KPP-----GRGKMHSSPVGRPSI-----GSASSSVVASGSSKEGSPLHKQDAGPS 53	54 SAGATT----TVSEKPGPKAAEVGDDFLGDBFVVWSERVWYNGVKPSPVVOYLETQAPGQW 109	
Db	672 SKVFAEPFKLIGQNHAKDSQYLVLPVIVD-DWLGGAAATSQHILLERMRSAQVTFAKAP 730	Qy 362 SKSKRNTWYIDEVAEDPAKSLEISTDFDR-----HSLPPSLTKP--NTNGSIGHSPPLSISAQS 437	362 SKSKRNTWYIDEVAEDPAKSLEISTDFDR-----HSLPPSLTKP--NTNGSIGHSPPLSISAQS 437	
Qy	744 CLIJQMPFRGKDPKLFKKIFPSLSENLNTLLETPQRCRIGGLAMTCRECYDDPDISA 803	Db 54 SAGATT----TVSEKPGPKAAEVGDDFLGDBFVVWSERVWYNGVKPSPVVOYLETQAPGQW 109	54 SAGATT----TVSEKPGPKAAEVGDDFLGDBFVVWSERVWYNGVKPSPVVOYLETQAPGQW 109	
Db	731 VLMQQLPQYQQ-KVFDKLPLPLETDITPFVAGVAPACKSICOACSEVYCTPCFLTRRYF 789	Qy 438 VMEELINTAPQESPLL-----AMPPQNSHGLEVGSSLAEVKENPPFY----478	110 AGVVLDDDPVGKNDGAVGGLRYYFECPALQIFITRISKLTROPAAESGSDGHVSITESITAQN 169	
Qy	804 GKKQFCKTCNTQVHLPKRLNHKYN---PVSLPKDLPDPWDWRHGCTPCQNNELFALVCI 860	Db 170 LSLSGTA---TPLTGRVIPRLREVSINNSVKRGNESGSNLSDGSVXRGDKDQJHLGDR 225	Qy 438 VMEELINTAPQESPLL-----AMPPQNSHGLEVGSSLAEVKENPPFY----478	410 -----KSCRPDRFA-----SLOPVSNQIERCNSLA-FGGYLSEVVBENTP 569
Qy	790 SEV-IFCRKCFHHTHLLPEIEDIKSRDLYPPGRPKQKPH-----SHRMVLSATLC 839	Db 226 VLYGGTKTGIVVRYGETDFAKGEMGVEL-DEPLGKNDGAVGATRYFQCCPKFGJFAPIH 284	Db 170 LSLSGTA---TPLTGRVIPRLREVSINNSVKRGNESGSNLSDGSVXRGDKDQJHLGDR 225	410 -----KSCRPDRFA-----SLOPVSNQIERCNSLA-FGGYLSEVVBENTP 569
Db	731 VLMQQLPQYQQ-KVFDKLPLPLETDITPFVAGVAPACKSICOACSEVYCTPCFLTRRYF 789	Qy 530 -----KSCRPDRFA-----SLOPVSNQIERCNSLA-FGGYLSEVVBENTP 569	Db 226 VLYGGTKTGIVVRYGETDFAKGEMGVEL-DEPLGKNDGAVGATRYFQCCPKFGJFAPIH 284	410 -----KSCRPDRFA-----SLOPVSNQIERCNSLA-FGGYLSEVVBENTP 569
Qy	861 ETSHYVAFVKYGRKDDSAMIFFDMSADRGONGFNPIPOVTPCBEGYELKMS---LEDL 916	Qy 570 PKMEKEGLIMIGKKGQIOTHYNCSYLDSTLFCFLFAFSVLDTVLRLP--KEKND-VEYY 626	Db 285 KVIRGFFSTSPAKAKKTKRMANGVSLTHSPSSSISVSSVASSVSGRPS-----336	410 -----KSCRPDRFA-----SLOPVSNQIERCNSLA-FGGYLSEVVBENTP 569
Db	840 ETSHYVAVTR--TSSNQVWFFDSMADRGQLSDGENVPVRECGRNSMDWLSQQWNRLKDA 897	Db 337 -----RSGLLTETSSRYA-----RKISGTATQEAALKOOGHIEQL 372	Db 285 KVIRGFFSTSPAKAKKTKRMANGVSLTHSPSSSISVSSVASSVSGRPS-----336	410 -----KSCRPDRFA-----SLOPVSNQIERCNSLA-FGGYLSEVVBENTP 569
Qy	917 H-----SLDSRRIQ-----GCARRLLCDAYMCMYQ--SPTMS 946	Qy 686 IL--RVEPLLKIRSAGQKVQDCYFYQ1FMVKNEKVGVPTIQQL-----LEWSPINSNL 736	Db 627 SETQELLRTIEVNPLRIGYVCATKIMKLRLKLEVKREA-SGFTSEKOPPEFLNLFH 685	410 -----KSCRPDRFA-----SLOPVSNQIERCNSLA-FGGYLSEVVBENTP 569
Db	898 DECGQIKVWELNRNISPNDRKLEIAMFGKQSSLDPVGRLLSDSYCIFYEDASPTSS 954	Db 373 LAERDLEPAREVAK-----ATSHICEKEITALKQAEHQVYAAEEKLQRARLL 424	Db 425 VRKEVKDLSNQLEEEERRKVEDLQF--RVEEESTIKGGDLETQTLHARIGELEQSIIL--L 480	410 -----KSCRPDRFA-----SLOPVSNQIERCNSLA-FGGYLSEVVBENTP 569
RESULT 15			Qy 686 IL--RVEPLLKIRSAGQKVQDCYFYQ1FMVKNEKVGVPTIQQL-----LEWSPINSNL 736	410 -----KSCRPDRFA-----SLOPVSNQIERCNSLA-FGGYLSEVVBENTP 569
ID	055156 PRELIMINARY;	PRT; 1046 AA.	Db 425 VRKEVKDLSNQLEEEERRKVEDLQF--RVEEESTIKGGDLETQTLHARIGELEQSIIL--L 480	410 -----KSCRPDRFA-----SLOPVSNQIERCNSLA-FGGYLSEVVBENTP 569
AC	055156;		Qy 737 KFAEAPSC1 745	410 -----KSCRPDRFA-----SLOPVSNQIERCNSLA-FGGYLSEVVBENTP 569
DT	01-JUN-1998 (TREMBLrel. 06, Created)		Db 481 EKAQAEERLL 489	410 -----KSCRPDRFA-----SLOPVSNQIERCNSLA-FGGYLSEVVBENTP 569
DT	01-JUN-2003 (TREMBLrel. 06, Last sequence update)			
DT	01-JUN-2003 (TREMBLrel. 24, Last annotation update)			
DE	CLIP-115 protein.			
GN	Rattus norvegicus (Rat).			
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			

Search completed: September 17, 2004, 09:48:16
 Job time : 148 secs

Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_p2n model

Run on: September 18, 2004, 15:16:48 ; Search time 8073 Seconds
(without alignments)

5095.075 Million cell updates/sec

Title: US-09-671-687A-3

Perfect score: 5034

Sequence: 1 MSSGLWSQEKVTPSYWEERI.....RLLCDAYMCMYQSPPMSLYK 949

Scoring table: BLOSUM62

Xgapop	10.0	Xgapext	0.5
Ygapop	10.0	Ygapext	0.5
Egapop	6.0	Egapext	7.0
Delop	6.0	Delext	7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing First 45 summaries

Command line parameters:

```
-MODEL=frame_p2n.model -DEV=xLP
-Q-/cgn2.1/OSPTO/spool/p/vn9671687/runat_17092004_103034_2781/app_query.fasta_1..1095
-DB=GenImbl -QFMT=fastaD -SUFFIX=large -MINWATCH=0..1 -LOOPCFL=0 -LOOPEXT=0
-UNITS=BITS -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40_cdi -LIST=45
-DCALIGN=N=200 -PCT=50 -THR=MAX=100 -THR MIN=0 -ALIGNN=15 -MODE=LOCAL
-OUTFMT=ptp -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US-09-671-687 @CGN_1..5932 @runat_17092004_103034_2781 -NCEPU=6 -ICPU=3
-NO_MMAPP -LARGEQUERY -NEG_SCORES=300 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -TREADS=1 -XGAPOP=10 -XGAPEXT=0..5 -FGAPPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0..5 -DELEXT=7
```

Database : GenEmb1:
1: gb_ba:
2: gb_htg:
3: gb_in:
4: gb_cm:
5: gb_ov:
6: gb_dat:
7: gb_ph:
8: gb_pl:
9: gb_pr:
10: gb_ro:
11: gb_sts:
12: gb_sy:
13: gb_un:
14: gb_vl:
15: em_ba:
16: em_fun:
17: em_hum:
18: em_in:
19: em_mu:
20: em_om:
21: em_or:
22: em_ov:
23: em_pat:
24: em_ph:
25: em_pi:
26: em_ro:
27: em_sts:
28: em_un:

29: em_vl:
30: em_htg_yum:
31: em_htg_inv:
32: em_htg_other:
33: em_htg_mus:
34: em_htg_pln:
35: em_htg_rdn:
36: em_htg_nam:
37: em_htg_vrt:
38: em_sy:
39: em_htgo_hum:
40: em_htgo_mus:
41: em_htgo_other:

* Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4983	99.0	3540	BC012342	BC012342 Homo sapi
2	4983	99.0	5414	AB020556	AB020556 Homo sapi
3	4971.5	98.8	5371	HSA250014	AJ250014 Homo sapi
4	4942	98.2	4527	BD231207	BD231207 Human cyt
5	4718.5	93.7	4314	AK122389	AK122389 Mus muscu
6	4718.5	93.7	4501	BC02438	BC02438 Mus muscu
7	4601	91.4	2845	AX883937	AX883937 Sequence
8	4601	91.4	2845	BD160617	BD160617 Primer fo
9	4601	91.4	2845	AK024348	AK024348 Homo sapi
10	4600.5	91.4	241990	AC123449	AC123449 Rattus no
C	11	4600.5	91.4	251122	AC020162 Rattus no
12	3676	73.0	2523	AR338799	AR338799 Sequence
13	3433.5	68.2	2341	AK000187	AK000187 Homo sapi
14	3022.5	60.0	2569	AX883681	AX883681 Sequence
15	3022.5	60.0	2569	BD160470	BD160470 Primer fo
16	3022.5	60.0	2569	AK024212	AK024212 Homo sapi
17	2930	58.2	1954	AK056226	AK056226 Homo sapi
18	2493	49.5	24163	AC131881	AC131881 Rattus no
19	2468	49.0	2116	AX012126	AX012126 Sequence
20	2468	49.0	2116	BD135433	BD135433 Receptor
C	21	2108.5	41.9	215717	AC133551 Mus muscu
22	1736	34.5	251957	AC126867	AC126867 Rattus no
23	1662	33.0	2426	BC049879	BC049879 Mus muscu
24	1442.5	28.7	212531	AC145178	AC145178 Gallus ga
25	1430.5	28.4	129025	AC145180	AC145180 Gallus ga
26	1262	25.1	837	AX863679	AX863679 Sequence
27	1262	25.1	837	BD149741	BD149741 Primer fo
28	1215	24.1	758	AX867014	AX867014 Sequence
29	1215	24.1	758	BD147076	BD147076 Primer fo
30	1132.5	22.5	2068	AY071592	AY071592 Drosophil
31	1111	22.1	151567	AC145503	AC145503 Canis fam
32	1111	22.1	199277	AC145445	AC145445 Canis fam
33	1109.5	22.0	208061	AC145254	AC145254 Bos tauri
C	34	1108.5	22.0	168271	AC007728 Homo sapi
35	1097.5	21.8	141663	AC145018	AC145018 Felis cat
36	1095.5	21.8	163319	HSA301140	HSA301140 Homo sapi
37	1085	21.6	155857	AC145174	AC145174 Pan trogl
38	1085	21.6	178790	AC145338	AC145338 Pan trogl
39	1078	21.4	194869	AC145003	AC145003 Papio anu
C	40	1077.5	21.4	188742	AC138025 Mus muscu
41	1067	21.2	232509	AC134093	AC134093 Rattus no
C	42	936	18.6	84367	AC005454 Drosophil
43	936	18.6	125011	AC020183 Drosophil	
C	44	936	18.6	153505	AC005849 Drosophil
45	936	18.6	154779	AC032237 Drosophil	

ALIGNMENTS

RESULT 1

BC012342	LOCUS	BC012342	mRNA	linear	PRI 04-OCT-2003		gene
	DEFINITION	Homo sapiens cylindromatosis (turban tumor syndrome), mRNA (cDNA)					/note="Vector: pORT7"
	Clone	MGC:19993 IMAGE:4552767					/note="Gene: CYLD"
	ACCESSION	BC012342					/note="Synonyms: CDMT, CYLD1, EAC, KIAA0849, HSPC057"
	VERSION	BC012342.1					/db_xref="LocusID:1540"
	KEYWORDS	MGC.					/db_xref="MIM:605018"
	SOURCE	Homo sapiens (human)					277 . 318
	ORGANISM	Homo sapiens					/codon_start=1
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;					/product="CYLD protein"
		Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.					/protein_id="AAH12442.1"
REFERENCE	1	(bases 1 to 3540)					/db_xref="GI:15214434"
AUTHORS	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Alischul, S.F., Zeeberg, B.R., Buetow, K.H., Schuler, G.D., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.E., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Cesavant, T.L., Schecter, T.B., Brownstein, M.J., Usdin, T.B., Arnbrink, P., Prange, C., Raha, S.S., Toshiyuki, S., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malik, J.A., Gunaratne, P.H., Richards, S., Worley, X.C., Hale, S., Garcia, A.M., Gay, L.J., Muzyk, S.W., Villalon, D.K., Muzyk, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahay, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, B.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J.J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalius, D.E., Schnarch, A., Schein, J.E., Jones, S.J. and Marra, M.A.						
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences						745 . 867
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)						/note="CAP_GLY: Region: CAP-GLY domain. CAP stands for cytoskeleton-associated proteins"
MEDLINE	22388257						/db_xref="CDD:pfam01302"
PUBMED	12477932						1720 . 2025
REFERENCE	2	(bases 1 to 3540)					/note="NIP100; Region: COG5244, NIP100, Dynactin complex subunit involved in mitotic spindle partitioning in anaphase B (Cell division and chromosome partitioning)"
AUTHORS	Strausberg, R.R.						/db_xref="CDD:COG5244"
TITLE	Submitted (15 AUG-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA						ORIGIN
JOURNAL	NIH-MGC Project URL: http://mgc.nci.nih.gov						Alignment Scores:
REMARK	Contact: MGC help desk						Pred. No.:
COMMENT	Email: cgapbs-r@mail.nih.gov						Score:
	Tissue Procurement: ATCC						Percent Similarity:
	CDNA Library Preparation: Rubin Laboratory						Best Local Similarity:
	Arrayed by: The I.M.A.G.E. Consortium (LNL)						Query Match:
	DNA Sequencing by: Genome Sequence Centre,						DB:
	BC Cancer Agency, Vancouver, BC, Canada						US-09-671-687A-3 (1-949) x BC012342 (1-3540)
	info@bcgsc.ca						
	Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Buttersfield, Sennia Chan, Readman Chiu, Chris Fjeil, Erin Garland, Ran Guin, Letricia Hsiao, Martin Krzywinski, Reta Kutche, Oliver Lee, Soon Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Sen, Pavani Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smailus, Michael Smith, Loraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasha van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.						
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium LNL at: http://image.lnl.gov						
	Series: IRAL, Plate: 29 Row: C Column: 5						
	This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 14165257.						
FEATURES	Location/Qualifiers						
Source	1 . 3540						
	/organism="Homo sapiens"						
	/mol_type="mRNA"						
	/db_xref="taxon:9606"						
	/clone="MGC:19923 IMAGE:4552767"						
	/tissue type="Uterus, leiomyosarcoma"						
	/clone_id="NIH_MGC_46"						

Db	577 TTGGCAATTACCAATTGAGGAGGGTTCAGCCCTGTTAACAAACAGAACATAAGT	636	Qy	477 PheTyrGlyvalleIarg9TrpIleGlyGlnProProGlyLeuAsnGluValleAlaGly	496
Qy	120 LysGlyLeuGlnIleAspValGlyCysProVallysvalGlnLeuArgSerGlyGluGlu	139	Db	1717 TTCTTGCGGTATATCGTTGATCGGTAGAAGTCAGCTGGAAAGA	1776
Db	637 AAAGCCCTCAAATAGACCTGGCTGTCGTGAAGTACAGTCAGCTGGAAAGA	696	Qy	497 LeuGluLeuGluAspGluCysAlaGlyCysThrAspGlyLysThrPheArgGlyThrArgY	516
Qy	140 LysPheProGlyValValArgPheArgGlyProleuLeuAlaGluArgThrValSerGly	159	Db	1777 CTGGAACCTGAGATGAGTGTGAGCTGAGGACTGAGGACTCTCGTAT	1836
Db	697 AAATTCCTGAGTGTAACTTCAGAACGCCCTGAGAGGAGACATCCGGA	756	Qy	517 PheThrCysAlaLeuIlysSylSalalenePheVallysLeuAsnSerCysAspProAspSer	536
Qy	160 IlePhePheIcyValGluLeuLeuGluGluGlyArgGlyClnGlyPheThrAspGlyVal	179	Db	1837 TTACACCTGTCGCTTCAGCAGCTTTCAGCTGACGGGCGC	1896
Db	757 ATATCTCTTGGTAGTGTAACTGCTGAAAGGGTCCTGTCAGCTTCACTGACGGGCGC	816	Qy	537 ArgPheAlaSerIleGlnProValSerAspGlnIleGluLrgCysAsnSerLeuAlaPhe	556
Qy	180 TyrGlnGlyLysGlnLeuPheGlnCysAspGluAspCysGly--pheValAlaLeuAsp	198	Db	1897 AGGTTGCATCATTCAGCAGCGGGTTCAATCAGATGAGCTGAGGCTGTGACTCTTGTGATT	1956
Db	817 TACCAAGGAAACAGCTTTTCAGTGTGATGAAATTGTCGCGMCTTGTGATTTGAC	876	Qy	557 GlyGlyTyrLeuSerGluValGluGluAsnThrProProlyAsmetGluLySgluGly	576
Qy	199 LysLeuGluLeuIleGluAspAspThrAlaLeuGluSerAspTyrAlaGlyProGly	218	Db	1957 GGAGCTACTAAAGTGAAGTAGTAGTGAAGAAATGAAAGAAATGAAAGAAGGC	2016
Db	877 AAGCTGAACTCATAGAAAGTGTGAAAGTGTGATGACATCGATTGAAAGTGTGAACT	936	Qy	577 LeuGluIleMetIleGlyLysLysIleGlyIleGlnGlyIleIstYAsnSeCysSerLeu	596
Qy	219 AspThrMetGlnValGluLeuProProLeuGluIleAspSerArgValSerLeuIlysGly	238	Db	2017 TTGGAGATATGATTGGAGAAAGAAAGGATCAGGGTCATTACAAATTCTGTGACTTA	2076
Db	937 GACAATTCATCGGTCGAACTTCCTCCCTTCCTGAAATAAACCTCCAGAGTTCTTGAAGGT	996	Qy	597 AspSerThrIlePheCysIleLeuPheAlaPheAlaPheSerSerValLeuAspThrValLeuLeuArg	616
Qy	239 GlyGlutThrIleGluSerGlyIleValIlePheCysAspValLeuProDlyLysGluSer	258	Db	2077 GACTGACCTTATTCGCTATTGCTTGTGACTGCTTACTGCTTACTTGTAGA	2136
Db	997 GGAGAACATAGATAGATCAGAACGTTATTCTGTGAGTTGTCAGGAAAGAAAGC	1056	Qy	617 ProlyGlyLysAspPheValGluIleArgThrSerGluLeuLeuArgThrGlu	636
Qy	259 LeuGlyTyrPheValGlyValAspMetAspAsnProIleGlyAlaTrraspGlyArgPhe	278	Db	2137 CCAAGAAAGAAACGATGAAATTATAGTGAACCCAAAGAGCTACTGAGCACAA	2196
Db	1057 TTAGATATTGTTGTTGTTGACATGGATAACCTTATGGCAACTGGAACTGGAGATT	1116	Qy	637 IleValAsnProLeuArgIleIteTyrolyTyValCysAlaThrIleMetLysLeuArg	656
Qy	279 AspGlyVal--LeuCysSerPheAlaCysValGluSerThrIleLeuLeuHisIleAsn	297	Db	2197 ATTGTAACTCTCTCGAAATATATGCTATGCTGTCGCAAAATAATTGAAACTGAGG	2256
Db	1111 7ATGGAGTCAGCTTGTGTTGCTGTTGTTGAAAGTACATTCTATGCACTCAT	1176	Qy	657 LysIleLeuGluLysValGluAlaIleSerGlyIlePheThrSerGluGluLysAspProGlu	676
Qy	298 AspIleLeProGluSerValIthrGlnGluIargArgProProlyLeuMetSer	317	Db	2257 AAATACTGAAANGTGGAGGGTCICATCAGGATTTACCTCTGAGAAAAGATCTGTGAG	2316
Db	1117 GATATCATCCAGAGGTGACGGAGTGTGACGGAGAAAGGGCCCTCCAAACTTCCTTATGTC	1236	Qy	677 GluPheLeuAsnIleLeuPheHisIleLeuArgValGluIleProLeuLeuIlysIleArg	696
Qy	318 ArgGlyValGlyAspIleGlySerSerHisasnLysProlysAlaThrGlySerThr	337	Db	2317 GAATTCTGATAATTCTGTTCATATATTAAAGGTGAAACCTTTGCTAAAATAAGA	2376
Db	1237 AGACGTGTTGGGGCAAAAGCTTCAGTCATATAAACCAAAAGGTPACAGGATCTAC	1296	Qy	697 SerAlaG1GlnLysValGlnNasPheCysTyrPheThrGlnIlePheMetGluLysAsnGlu	716
Qy	338 SerAspProGlyIasnArg--ArgSerGluLeuPheIleAspSerIleSerVal	356	Db	2377 TCAGGAGGTCAAACGTCAGATGTTACTCTPATCAATTTTTATGGAAAAAAATGAG	2436
Db	1297 TCAGACCTCTGGAAATAGAACAGATCTGAAATTATACCTTAAATGGCTTCCTGTT	1356	Qy	717 LysValGlyIalProThrIleGlnIleLeuGluTrpSerPheIleAsnSerAsnLeu	736
Qy	357 AspSerGlnProGlnSerIleSerSerHisasnLysProlysAlaThrGlySerThr	376	Db	2437 AAATGTGGCTTCCACAACTCAGAGTTTATCAGTGGCTTTATCAGTAACTG	2495
Db	1357 GACTCACACAACTCAAAATACAGTAAATTGTAATTGTAATTGTAATTGTAATTG	1416	Qy	737 LysPheAlaCluAlaProSerCysIleLeuGluIleGlnMetProAspPheGlyLysAspPhe	756
Qy	377 ProAlaLysSerIleThrGluIleSerThrAspHeAspArgSerSerProProLeuI	396	Db	2497 AAATTGACGGACCATCATGTCGTTGAAAGAACCTTGAATTTGAAACTTT	2556
Db	1417 CCTGCAAATTCTTACAGAGATCTACAGTCATCTACGCTTACACCAACTCCG	1476	Qy	757 LysIleLeuPheIleIlysIleLeuAsnIleThrAspIleLeuGluIAsp	776
Qy	397 ProProvalAsnSerThrAspIleLeuProLeuIleSerIleSerLeuIleSer	416	Db	2557 AAACATTTTAAATAAAATTTCCCTCTCGAAATTAAACAGATTACTGAGAC	2616
Db	1477 CCTCTCTCTGTAACCTACTGACCTACGCTTACATTCACCTTCAGTC	1536	Qy	797 AspAspProAspIleSerAlaGlyIlysIleIlysGlnPheCysIleThrCysAsnThrGln	816
Qy	417 ThrllysMetProAsnThrAspIleIlysIleSerIleSerLeuIleSerIleSer	436	Db	2677 GACGATCCGGACATTCAGCTGGAAATAACGATTTGTAACACTGAAACTCAA	2736
Db	1537 ACCAGATCCTCAATTACCAATGGAAGTGTGCACTTCAGCCAG	1596	Qy	817 ValHisLeuHisProLysArgLeuAsnHisLysTyrAsnProValSerLeuProlysAsp	836
Qy	457 ProGlyAsnSerHisGlyLeuGluValGlySerLeuIagluVallysCysAsnProPro	476	Db	2737 GTCCCACTTCATCCGAAAGGGCTCATATAACCCAGTGTGAACTTCCAAAGAC	2796
Db	1657 CCTGGAAACTCATGGTCATGGTCAATTGAGTGGCTCATGGTGAAGTTAAGGAGAACCTCT	1716			

Db	1352	TITATGTCAAAGAGGGTTGGGACAAGGGTCATCCAGTCATAAAACCAAAAGCTACA	1411	
Qy	335	GlySerThrSerAspProGlyAspArg--ArgSerGluLeuPheThrLeuAsnGly	353	
Db	1412	GGATCTACCTCAGACCCTGGAAATTAGAACAGATCTGAATTATTTATCCCTAAATGG	1471	
Qy	354	SerValAspSerGlnProGlnSerLysSerLysAsnThrTrpThrileAspGluVal	373	
Db	1472	TCTCTGTACTCACACCAATCAAAATCACGAGTACATTGTAAGTGAAGT	1531	
Qy	374	AlaGluAspProAlaSerLeuThrGluIleSerThrAspHeAspArgSerSerPro	393	
Db	1532	GCAGAGACCTGCAAATCTCTTACAGAGATACTACAGACTTGACCGTTCTTCACCA	1591	
Qy	394	ProLeuGlnProProValAsnSerLeuThrGluAsnArgPheIleSerLeuPro	413	
Db	1592	CCACTCCAGCCTCTCCCTGTGAACTCACTGACCCGAGAACAGATTCCACTCTTACCA	1651	
Qy	414	PheserLeuThrlysSerProAsnThrAsnGlySerIleGlyHisSerProLeuSerIeu	433	
Db	1652	TTCAGTCRACCAAGATGCCAAATACCAAGTCAAGTCACATTCCTCG	1711	
Qy	434	SerAlaGlnSerValMetGluGluLeuAsnThrAlaProValGlnGluSerProProLeu	453	
Db	1712	TCACCCCACACTCTGAAATGGAGACCTAAACACTCACCCTGACAGAGTGTCCACCCCTTG	1771	
Qy	454	AlaMetProProGlyAsnSerHisGlyLeuGluValGlySerLeuAlaGluValGly	473	
Db	1772	GCCATGCCCTCTGGAACTCACATGGCTAGAAGTGGGTCAATGGCTGAAGTAAAGG	1831	
Qy	474	AsnProProPheThrGlyValLeuArgTrpIleGlyGlnProProGlyLeuAsnGluVal	493	
Db	1832	AACCTCCCTTCATGGGTTAATCGGTTGATCGTCACCAAGGACTGATGAAGT	1891	
Qy	494	LeuAlaGlyLeuGluLeuGluAspGluCysAlaGlyCysSerGlyThrPheArgGly	513	
Db	1892	CTCCCTGGACTGGAACTGGAAAGATGAGTCGTCAGGCTGTACGGATGTGGAACTTCAGGGC	1951	
Qy	514	ThrArgGlyPheThrCysAlaLeuLysLysAlaLeuPheVallyLeuLysSerCysArg	533	
Db	1952	ACTCGGTATTCACCTGTCCTCTAGGGCTCTGAACCTGAAACTGAGAAGCTGCAGG	2011	
Qy	534	ProAspSerArgPheAlaSerLeuGlnProValSerAsnGlnIleGluGlyArgCysAsnSer	553	
Db	2012	CCTGACTCAAGTTGACCATTCAGGGTTCCATCAGATTGAGCTTAACCT	2071	
Qy	554	LeuAlaPheGlyTyrIleSerGluLeuIvalGluGluAsnThrProProLysMetGlu	573	
Db	2072	TTAGCATTTGGGGCTACTTAAGTAGTAGTAGAGAAATACTCCACAAAATGGAA	2131	
Qy	574	LysGluGlyLeuGluIleMetGlyLysLysGlyIleGlnGlyHistYrAsnSer	593	
Db	2132	AAGAAGGGTTGAGATATGTTGGAGAAAGAAAGAACGAGTAGATAATTGAGCTTAACATCT	2191	
Qy	594	CysteIleAspSerThrLeuPheCysLeuPheSerValleAspThrVal	613	RESULT 4
Db	2192	TGTGACTTAAGCTCAACCTTATCTGCTTATTGCTTGTGCTGACACTTG	2251	BD231207
Qy	614	LeuLeuArgProLysGluLysSerAspValGlyLysGlyIleGlnGlyHistYrAsnSer	633	LOCUS Human cytoskeleton associated proteins.
Db	2252	TCTACTGAGCCAAAGAAAGAACGAGTAGATAATTGAGCTTAACATCT	2311	DEFINITION BD231207 Human cytoskeleton associated proteins.
Qy	634	ArgThrGluIleValAspProLeuArgIleTyrGlyTyrValCysAlaArgGly	653	REFERENCE 1 (bases 1 to 40977)
Db	2312	AGGACAGAAATTGTTAATCTCTGAGAAATATGATGATGCTGCTGACACTTG	2371	AUTHORS Lal, P., Tang, T.Y., Yue, H., Hillman, J.L., Bandman, O., Corley, N.C., Guegler, K.J., Patterson, C., Azimzai, Y., and Baughn, M.R.
Qy	654	LysLeuArgLysIleLeuGluIysSerGluAlaSerGlyPheThrSerGluGluLys	673	TITLE Human cytoskeleton associated proteins.
Db	2372	AAACTGAGCAAAATACTGAAAGTTGAGGTGAAAGGCTGACACTTGAGAA	2431	JOURNAL Patent: JP 200526016-A 9 20-Aug-2002; INCYTE PHARMACEUTICALS INC
Qy	674	AspProGluGluPheLeuAsnIleLeuPheHisIleLeuArgValGluProLeuLeu	693	COMMENT OS Homo sapiens (human)
Db	2432	GATCTGAGGAATCTGAAATATCTGTTCAATATTAGGTAGAACCTTGGTA	2491	PN JP 200256076-A 9 20-Aug-2002
			PD 20-Aug-2002	

SIGNIN

Db	2195	CCCCAAAGAAAGATGATGATAAGATTACAGTGAGCTCAGGCTACTGAGCACAG	2254		BC042438.1	GI:27503670	
Qy	637	IleValAsnProLeuArgIleTyrGlyTyrValCysAlaThrlysIleMetlysLeuIg	656	MUS_musculus (house mouse)			
Db	2255	ATAGTCATCTCTGAGATAATGGATGTTGCAAGAAGATTGAACTGAGG	2314	MUS_musculus			
Qy	657	LysIleLeuGluLysValGluAlaAlaSerGlyIleThrsSerGluGluLysAspProGlu	676	ORGANISM			
Db	2315	AAAATACTGAAAGTGTAGGGCTGCATGGTTTACCTCTGAAGAAAAGTGCTCGA	2374	REFERENCE			
Qy	677	GluPhelLeuAsnIleLeuPheIshisIleLeuArgValGluProLeuLeuIysIleArg	696	AUTHORS	1 (bases 1 to 4501)		
Db	2375	GAATTCTTAATAATATCCCTGTTCATGATATTAAAGGTGAAACATTGATAATAAGA	2434	Klausner,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Schuler,G.J., Altschul,S.F., Collins,F.S., Wagner,L., Shenmen,C.M., Blat,N.K., Diatchenko,L., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Iocque,L., Elano,N.A., Peters,G.J., Abramson,R.D., Mullaney,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Gibbs,R.A., Fahey,J., Helton,B., Kettman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,B.D., Dickson,M.C., Grimwood,J.C., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schneroch,A., Schein,J.E., Jones,S.J. and Marrs,M.A.			
Qy	697	SerAlaGlyGlnIleGluAspCysTrpPheArgTrpSerPheIleAsnSerAsnIle	716	TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		
Db	2435	TGAGCAGTCAAAGTCAGACTGTTAACATTCTATCAAATTTTATGGAAAAAAATCG	2494	JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)		
Qy	717	LysValGlyValProThrIleGlnGlnIleLeuGluItrpSerPheIleAsnSerAsnIle	736	MEDLINE	22388257		
Db	2495	AAAGTGGAGTACCCAGATTCCGAGTATTAGAATGGCTTTTATCAACGCAACCTG	2554	PUBMED	12477932		
Qy	737	LysPheAlaGluAlaProSerCysLeuIleLeuGlnMetProArgPheGlyIysAsphe	756	REFERENCE	2 (bases 1 to 4501)		
Db	2555	AAATTGGAGGCCACCATCATGCTGATTATCCAGATGCCCTGGTTGGAAAGACTT	2614	AUTHORS	Direct Submission		
Qy	757	LysLeuPheIysIleLePheProSerIeuGlulenAsnIleThrAspIeuGluAsp	776	JOURNAL	Submitted (02-JAN-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
Db	2615	AACTATTAAGAAATTTTCCCTCCCTGGATTAAATATAACAGATTACTTGAAAC	2674	REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov		
Qy	777	ThrProArgGlnCysArgIleCysGlyLeuAlaMetTyrGluCysBrgGluCysTyr	796	COMMENT	Contact: MGC help desk Email: cgabbs@mail.nih.gov		
Db	2675	ACTCCCAGCAGTGGCGGATCTGTGGAGGACTGGATGTTACAGTGAGGGAGTGCTA	2734	TISSUE	Tissue Procurement: Gilbert Smith, Ph.D.		
Qy	797	ABDAspProAspIleSerAlaGlyLysIleGlnPheCysIlySthrCysAsnThrCln	816	JOURNAL	CDNA Library Preparation: Life Technologies, Inc.		
Db	2735	GACGATCGGAGCATCTAGCTGGAAAGATCAAGCTGTAAGACCTGCGACACTCG	2794	COMMENT	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILN) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center		
Qy	817	ValHisLeuIleIspolyArgLeuAsnHisLysIleTyrAspProValSerIeuProlAsp	836	CONTACT	Center code: BCH-HGSC Web site: http://www.hgsc.bcm.tmc.edu/cdna/		
Db	2795	GTTCACCTTCATCCGGAGGTGAATCATCCAGTATCATCCAGTATCTCTCA	2854	FEATURES	Contact: amg@bcm.tmc.edu		
Qy	837	LeuProAspTrpAspPTPArgIleGlyCysIleProCysGlnAsnMetGluLeuPheHeal	856	source	Gene Consortium/LILN at: http://image.liln.gov		
Db	2855	TGGCTTGACTGGGACTCTGGAGCTGGAGCTGGCTGTCATCCCTGTCAGAGGGTTGCT	2914	LOCATIONS	Series: IRAK Plate: 30 Row: 9 Column: 3.		
Qy	857	VallLeuCysIleIspolyArgLeuAlaPheValLysTyrGlyIysAspSer	876	QUALIFIERS	Location/Qualifiers		
Db	2915	GTCCTCTCTCATGAAACCAGGCCACTATGTCGAAAGTAGTCGAGGTGACTCT	2974		1..4501 "organism="Mus musculus"		
Qy	877	AlaTrpLeuPhePheAspSerMetAlaAspPargAspGlyGlyIysAspSerIeuLeuGluAspIeu	896		/mol type="mRNA"		
Db	2975	GCCTGGCTTCTTGACAGCATGGCGCATGGCTGAGTCGATGGCTCAACAT	3034		/strain="FVB/N"		
Qy	897	ProGlnValIleThrProCysProGluvalIgIysIleMetSerLeuGluAspIeu	916		/db_xref="Caxon:1090"		
Db	3035	CCACRAGTGACGCCAGTGAGTGAGACTGAGTGCTCTGAGGACCTG	3094		/clone="MGC:25429"		
Qy	917	HisSerLeuAspSerArgArgIleGlnGlyCysAlaArgLeuCysAspAlaTyr	936		tissue_type="Mammary tumor"		
Db	3095	CACTCTTGACTCCAGAGGATTCAACGCTGCGCCAGACTCTTGCATGCTACA	3154		model_10 month old virgin mouse.		
Qy	937	MetCYSMetTyrGlnSerProThrMetSerLeuTyrls	949		Location/lab_hos="Virus ID:74256"		
Db	3155	ATGTGCATGTACCAAGTCACCATGHCCTGTCACAA	3193		/note="vector: pcMV-SPORT6"		
RESULT	6			1..4501			
BC042438				/gene="CylD"			
LOCUS	BC042438	4501 bp	mRNA	/note= "synonyms: C130039D01Rik, 2900009M21Rik, EAC, CDMT, CYLD1"			
DEFINITION		Mus musculus cylindromatosis (turban tumor syndrome)	mRNA (cDNA Clone MGCI-25429 IMAGE:3983771), complete cds.	/db_xref="LocusID:74256"			
ACCESSION	BC042438			/db_xref="MGI:1921506"			

Db	Best Local Similarity:		99.32%
	Query Match:	DB:	91.40%
2222 TTTGTGAAACCTGCACACTCAAGTCRCCCTCATCGAGGGCTGATCATAATAAT 2281	Mismatches:	2	
829 AsnProValSerLeuProLysAspLeuProAspTyrAspSerArgHisGlyCysIlePro 848	Indels:	4	
2282 AACCGAGTCACTCCAAAGACTTACCGACTGGACTGGAGAACGGTGCATCCCT 2341	Gaps:	4	
US-09-671-687A-3 (1-949) x BD160617 (1-2845)			
849 CysGlnAsnMetGluLeuIleCysIleGluThrSerHisTyrValAlaPhe 868	Qy	73 LysIleLeuGluGlnProHisAlaValLeuPheValAspGlu - -AspValValGluIle 91	
2342 TGCCAAATATGGAGTTATTGCTCTCTCATAGAAACAAGGCCATCTGCTT 2401	Db	2 AAATCTAGAGCAACCTCATGCACTTCCTTGATGAAAGATGTCAGAGATA 61	
869 ValIleTyrGlyLysAspAspSerAlaIlePheAspSerMetAlaAspArgAsp 888	Qy	92 AsnGluLysPheThrGluLeuLeuAlaIleThrAsnCysGluGluArgPheSerIeu 111	
2402 GTGAGATGTGGAGGGACGATTCGCTGGCTCTGAGCTGGCGATGGAT 2461	Db	62 ATGAAAGCTCACAGTACTTGGCAATTACCAATITGGGATTACAGTGGAGGGTCAGCTG 121	
889 GlyGlyGlnAsnGlyPheAsnIleProGlnValThrProGlyCysProValoYs 908	Qy	112 PhenylAsnArgLeuSerIlysGlyLeuGlnIleAspValGlyCysProValoYs 131	
2462 GGTGGTCAGATGGCTCACATTCTCAAGTCAACATTGCAAGTAGGAGGTAC 2521	Db	122 TTAAAACAGAAAACAGAAACAGACTAAGTAAGGCCTCCAATAAGCTGTGTGAA 181	
909 LeuIysMetSerIeuGluAspLeuHisSerLeuAspSerIrgArgIleGlnGlyCysAla 928	Qy	132 ValGlnIleArgSerGlyLysGluGlnIleAspProGlyValValArgPheArgClyProLeu 151	
2522 TTGAGATGCTCTGGAAAGACCTGCAATTCTGGACTTCCAGGAGATCCAAAGGCTGTGCA 2581	Db	182 GTACAGCTGAGATGGGAAGAAAATTTCTGGAGATTACGGCTTCAGGGACCCCTG 241	
929 ArgArgLeuIleCysAspAlaTyMtyMetCysMetTyrGlnSerProThrMetSerIeuTyr 948	Qy	152 LeuAlaGluArgThrValSerGlyIlePhePheGlyValGluLeuIleGluClyGluArg 171	
2582 CGAGAGCTGCTTGTGATGCAATATGCTGATGCAATGCTAACATGAGTTGTAC 2641	Db	242 TAGCAGAGGGAGCTGTCGGGAAATCTTGGAGTGAAGGAGGAGGCTGGAGGCTGGGT 301	
949 Lys 949	Qy	172 GlyGlyLysPheHisAspGlyValYtGlyLysGlnIlePheGlnCysAspGluAsp 191	
Db	302 GGTCAAGGTTCACTGACGGGTCTGGAAACAGCTTTCAGTGTGATGAGAT 361		
RESULT 8	Qy	192 CysGly -- PheValAlaIleAspIlysLeuGluLeuIleGluAspAspThrAlaLeu 210	
BD160617	Db	362 TGCGCCGTTGTTGTTGCATTGGACAGCTGAACTCATAGAGATGACATGCTATTG 421	
DEFINITION Primer for synthesizing full-length cDNA and use thereof.	Qy	211 GluUserAspTyrAlaGlyProGlyAspThrMetGlnValGluLeuProProleuGluIle 230	
ACCESSION BD160617	Db	422 GAAGGTGATTACGGGCTCTGGAAACACATGGGTGAACTCTCTCTGGAGATA 481	
VERSION 1.1	Qy	231 AsnSerArgValSerLeuLysGlyGlyGluThrIleGluUserGlyIlyThrValleHeCys 250	
KEYWORDS Homo sapiens (human)	Db	482 AACTCAGAGTTCTTGAAGGTTGAGCTTCTGAGACAGTTATCTGGAAACATAAGAACT 541	
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	Qy	251 AspValLeuProGlyLysGluSerIeuglyTyrPheValGlyValAspMetAspAsnPro 270	
REFERENCE 1. (bases 1 to 2845)	Db	542 GATGTTTGGCCAGGAAAGAAAGTTAGGATATTTGTTGGTGTGAGCATGATAACCT 601	
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J., Ishii,S., Sugiyama,T., Wakamatsu,A., Naga,K. and Obsuki,T.	Qy	271 IleGlyAsnTrpAspGlyArgPheAspGlyVal --LeuCysSerPheAlaCysValGlu 289	
SOURCE Homo sapiens (human)	Db	602 ATGGCACTGGATGGAGATTGATGAGCTTGTGTTGCTGTGAA 661	
VERSION JP 2002191363-A/154660.	Qy	290 SerThrIleLeuIleAsnAspIleLeuPheGluSerIleGluGluArg 309	
PRIMER Primer for synthesizing full-length cDNA and use thereof	Db	662 AGTACAATTCTATTGACATGATACTCCAGAGGTGTGACGACAAAGGAGG 721	
JOURNAL Patent: JP 2002191363-A 15460 09-JUL-2002,	Qy	310 ProProlylSerLeuIlePheMetSerArgGlyValGlyAspGlySerSerIleAsn 329	
COMMENT OS HUMAN RESEARCH INSTITUTE	Db	722 CCTCCCAAACCTGGCTTATGTCAGGATCTCATGATACTCCAGAGGTGTGACGACAAAGGAGG 781	
PI KEIICHI NAGAI, TETSUJI OTSUKI	Qy	330 LysProlylSerLeuIlePheMetSerArgGlyValGlyAspGlySerSerIleAsn 348	
PC C12N15/09, C07K14/47, C07K16/18, C12N1/19, C12N1/21, C12N5/ PC	Db	782 AAACCAAGGTACAGGATCTACCTAGACCTGAAAGATCTGAATATT 841	
10, C12P21/02, C12Q1/68 // C12P21/08, G06F17/30, C12N5/00, C12N5/00 CC	Qy	349 TyrThrLeuAsnGlySerSerIleAsnGlySerSerIleAsnThrTyr 368	
PC Primer for synthesizing full-length cDNA and use thereof FH Key	Db	842 TATACCTTAATGGCTTCTGACTCACAAACACATCAAAATACATGG 901	
FT CDS Location/Qualifiers	Qy	369 TyrIleAspGluValAlaGluAspProAlaLysserLeuIleSerthrAspPhe 388	
FEATURES 1. .2845	Db	902 TACATGTAAGTGGAGAACCCCTCATCAGAGATATCTACAGACTTT 961	
source /organism="Homo sapiens"	Qy	389 AspArgSerSerProProLeuGlnProProLeuAsnSerLeuThrThrGluAsnArg 408	
/organism="Genomic DNA"			
/db_xref="taxon: 9606"			
ORIGIN Alignment Scores:	Qy		
Pred. No.: 0	Length:	2845	
Score: 99.32%	Matches:	875	
Percent Similarity: 0	Conservative:	0	

etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.
 Location/Qualifiers
 1 .2845 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="PLACE1006129"
 /tissue_type="placenta"
 /clone_lib="PLAC1"
 /note="Cloning vector: pME18SFL3"

ORIGIN

Alignment Scores:

Pred. No. :	0	Length:	2845	Db	722 CCTCCCCAAACTGCCCCTTATGCAAGGGTGTGGGACAAAGGTCATCCAGTCATAAT 781
Score:	4601.00	Matches:	875	Qy	330 LysProLysAlaThrGlySerThrSerAspProGlyAsnArg--ArgSerLeuLeuPhe 348
Percent Similarity:	99.32%	Conservative:	0	Db	782 AACCAAAGGTACAGGATCTACTCGACCTGGAAATGAAACAGATCTGAATTATT 841
Best Local Similarity:	99.32%	Mismatches:	2	Qy	349 TyrThrLeuAsnGlySerSerValAspSerGlnProGlnSerLysAsnThrTrp 368
Query Match:	91.40%	Indels:	4	Db	842 TATACCTTAATGGCTCTTGTTGACTCAGAACAAATCAAATCATGG 901
DB:	9	Gaps:	4	Qy	369 TyrIleAspGluValAlaGluAspProAlaLysSerLeuThrGluIleSerThrAspPhe 388
us-09-671-687A-3 (1-949) × AK024348 (1-2845)				Db	902 TACATGTGATGAACTTCAGAGACCTGCAAAATCTCTTACAGAAATCACAGACTCACAGACTTT 961
Qy	73 LysIleGluGluGlnProLysAlaValLeuAspLeuPheValAspGlu--AspValValGluIle 91			Qy	389 AspArgSerSerProProLeuGlnProProValAsnSerLeuThrThrGluAsnArg 408
Db	2 AAAATTCTGAGCACCCTATGCAGTTCTCTGTTGATGAAAGGATTTGAGAGATA 61			Db	962 GACCGTTCTTACACCACACTCAGGCTCCCTGGAACACTGACCTGAACTGACCTGCCCAA 1021
Qy	92 ArgGluLysPheThrGluLeuLeuAlaLeuThrAlaCysGluGluArgPheSerIeu 111			Qy	409 PheHisSerIeuProPheSerLeuThrLysMetProAsnThrAsnGlySerIleGlyHis 428
Db	62 AATGAAAGCTCACAGAGTACTTTGGAAATTACCAATTGTTGAGGAGGGTTCAGCCCTG 121			Db	1022 TTCCCACTCTTACATTCAAGCTCAAGTCAACAAATGCCCATAATGAAATGGCCAC 1081
Qy	112 PhoLysAsnArgLeuSerArgLeuSerLysGlyLeuGlnIleAspValGlyIcySpValIys 131			Qy	429 SerProLeuSerLeuSerAlaGlnSerValMetGluGluIleAsnThrAlaProValGln 448
Db	122 TTAAAAACAGAACAGCTTAAGTAAGAACCCCTCCAAATAAGTACCTGGGTTCTGTGAA 181			Db	1082 AGTCACACTTCCTCTCAGGCTCTGCTGAACTGAGGTTCTATGGAGCTCATCTG 1141
Qy	132 ValGlnLeuArgSerGlyGluGluLysPheProGlyValValArgPheArgGlyProLeu 151			Qy	449 GluSerProProLeuAlaMetProProGlyAsnSerSerHisGlyLeuGluValGlySerLeu 468
Db	182 GTACAGCTGAGATCTGGGGAGAAAATTCTGGAGTTGATGCTGAGCTTGAGAACCCCTG 241			Db	1142 GAGACTCCACCCCTTGCCATGCCCTGCTGGAAACTACATGCTGAACTGAGCTG 1201
Qy	152 LeuAlaGluArgThrValSerGlyIlePhePheGlyValGluIleGluGluGlyArg 171			Qy	469 AlaGluValLysGluGluAsnProProPheTerGlyValValLeuArgTerGlyGlnProPro 488
Db	242 TTAGCAGAGGGCAAGTCCTCCGGRATAATTCTTGGTGAATTGCTGGAAAGGGTCTG 301			Db	1202 GCTGAGATTAAGGAAACCCCTCCCTCTATGGGAAATCCTGCTGATGCTGAGCTG 1261
Qy	172 GlyGlnGlyPheThrAspGlyValValGlySerIlePheGlnGlyAspGluAsp 191			Qy	489 GlyLeuAsnGluValLeuGluGluAspGluCysAlaGlyCysThrAsp 508
Db	302 GGTCAAGTGTTCCTGAGGGTGTACCAAGGGAAACAGCTTTCTGTTGAGCTGAA 361			Db	1262 GGACTGAACTGAACTGCTGAGATGAGTGTGAACTGAGGTGACGGTGTACGGAT 1321
Qy	192 CysGly--PheValAlaLeuLeuAspLeuGluLeuGluAspAspThrAlaLeu 210			Qy	509 GlyThrPheArgGlyThrArgTrpThrCysAlaLeuIlePheValPheValLys 528
Db	362 TGTCGCTGTTGATGGCATGGAAAGCTAGATGAGATGAGACTGCAATTGCA 421			Db	1322 GGAACCTTCAGGGGACTCTGGTATTCTACCTGCTGAAAGGCGCTTTTGAA 1381
Qy	211 GluSerIspTyrAlaGlyProGlyAspThrMetGlnValGluLeuProProLeuGluIle 230			Qy	529 LeuIysSerCysAspGlyProAspSerArgPheAlaSerLeuGlnProValSerAsnGlnIle 548
Db	422 GAAAGTGATTAACGGCTCTGGGACACAATCAGGTGCAACTTGGTGAATACTGAA 481			Db	1382 CTGAGAGCTGACGCCCCTGACTCTGGTATTGCTGCTGATTTGCAATTCTCGATT 1441
Qy	231 AsnSerArgValSerLeuIleLeuGluLeuGluThrIleGluSerGlyThrValIlePheCys 250			Qy	549 GluArgCysAsnSerLeuAlaPheGlyGlyTYrLeuSerGluValGluAsnThr 568
Db	482 AACTCCAGTTCTGAGGTTCTGAGGTGCAAGCTTGGTGAATACTGAA 541			Db	1442 GAGCCCTGTAACCTCTTACCTTGAGCTACTAACTGAGTACTAGAGAAATACT 1501
Qy	251 AspValLeuProGlyLysGluSerLysLysPheAspValGlyValAspMetAspAsnPro 270			Qy	569 PropolymerGluLysGluGluIleMetIleGlyIleGluGluIleGlyIleGln 588
Db	542 GATGGTTGCCAGAAAGAAAGCTTAGGATATTGTTGGTGTGGAACTGGATAACCT 601			Db	1502 CCACAAAATGGAAAAGAGGCTGGAGATAATGATTGTTGGATTCAGCAAC 1561
Qy	271 11eGlyAsnTrpAspGlyArgPheAspClyIval--LeuCysSerPheAlaCysValGlu 289			Qy	589 GlyHisTerAsnSerCysIleSerThrLeuPheCysLeuPheAlaPheSerSer 608
Db	602 ATGGCAACTGGGATGGAAGTGTGAGTTGAGCTTGGTGTGGTGTGTTGTTG 661			Db	1562 GGTCTTACAAATTCTGCTTACTGACTGACCTTATTGCTGTTTTGCTTTCCTCT 1621
Qy	290 SerThrIleLeuHisIleLeuAsnAspIleLeuProGluSerValThrGlnGluArg 309			Qy	609 ValLeuAspThrValLeuLeuArgProLysGluIysAsnAspValGluTyrrSerGlu 628
Db	662 AGTACAATTCTATTGCACTCATGATCATGAACTGAGTGTGAGCTTGGTGTGG 721			Db	1622 GTRCTGGGACACTGTGTTACTTGGACCAAGAAAGAACATGTTGAAATTATAGCTGAA 1681
Qy	310 ProProLysLeuAlaPheMetSerArgClyValGlyAspLysGlySerSerHisAsn 329			Qy	649 AlanThrLysIleMetLysLeuIleLeuAsnSerGlyPhe 668
Db	1802 ACCCTGAAAGAAAGATCTGAGGAATCTGATATCTGTTCATATACTGTTAAGG 1861			Db	1742 GCCAAAGAAATTGATGTTGAACTTCTGAAATTGTTGAAATTGATGTTGTTGTT 1801

Qy	689	ValGluProLeuLeuIysSerIleArgSerAlaGlyGinLysValGinAspCysTyrPheTyr	708	Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Barnstead, M., Benahmed, F., Baldwin, D., Bandaranaike, D., Barber, M., Blankenburg, K., Blyth, P., Brown, M., Biswalo, K., Blair, J., Blaustein, P., Burrell, K., Calderon, B., Bryant, N., Buhay, C., Burch, K., Carter, K., Cavarozzi, I., Caesar, H., Center, A., Cardenas, V., Chacko, J., Chavez, D., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cochrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davis-Carroll, L., De Andrade, C., Dederich, D., Delgado, O., Denson, S., Derano, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escott, M., Eugene, C., Evans, C.A., Failes, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebara, G., Guevara, W., Guevara, W., Hernandez, K., Gunarate, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlik, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodges, A., Hollins, B., Howells, S., Hulk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., King, L., Kovar, C., Kovari, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lohse, R.J., Lozada, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindarne, M., Mahmood, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenin, E., Milosavljevic, C.A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munisoba, M., Murphy, M., Neir, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwakelume, O., Okwuonu, G., Olarinpungsoo, A., Parks, K., Paszternak, S., Paul, H., Perez, L., Perez-L., Prankoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puerto, M., Quiroz, J., Rachlin, P., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scheer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergran, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinle, M., Strong, R., Surton, A., Svarek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, L., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Williamson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J.-Y., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Smith, D.R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.	Unpublished 2 (bases 1 to 241990)
Db	1862	GTAGACCTTGCTAAATAAGTAGCATGGTCAGATGTTACTCTAT	1921		Direct Submission 2 (bases 1 to 241990)
Qy	709	GlnIlePhenMetGluIysSerAlaGluValProThrIleGlnLeuIeuGlu	728		Direct Submission 2 (bases 1 to 241990)
Db	1922	CAAATTGTTATGGAAAAAATGAGAAAGTCGCTCCCAATTCACTGACTGTAGAA	1981		Submitted (29-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
Qy	729	TrpSerPheIleAsparBleuIleSphAlaGluAlaProSerCysBleuIleIleGln	748		On Sep 21, 2002 this sequence version replaced gi:21903787. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using ATLAS (http://www.hgsc.bcm.edu/projects/rat/). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
Db	1982	TGGTCMTCATTACACGTAACTGAAATTGAGCCACATTGCTGATTTAC	2041		-----
Qy	749	MetProArgPheGlyLysAspBphelysLeuPhelysLysIlePheProSerLeuGluLeu	768		-----
Db	2042	ATGCCCTGATTGGAAAAGCTTAACATTAAACTTAAATAATTTCCTCTGGATT	2101		-----
Qy	769	AsnIleThrAspLeuGluAspSerThrProArgGlnCysBargIleCysGlyGlyLeuAla	788		-----
Db	2102	AATAAACAGATTACTTGAGAGACACTGAGACATCGCCGATATGTGGAGGGCTGCA	2161		-----
Qy	789	MecTyGluCysArgGluCystyramPaspProAspIlesSerAlaGlyLysIleGlyLysIleGln	808		-----
Db	2162	ATGTATGAGTGTAGAGAAATGCTACAGACATCCAGACACTGAGACATCGCCGATTCAGCTGAA	2221		-----
Qy	809	PheCysLysThrCysAsnTrnglnValHisIleHisProlysArgLeuAnhishistYR	828		-----
Db	2222	TTTGTGAAAAACCTGCAACATCAAGTCACCTTCATCGAGAGGCTGAATCATATAAT	2281		-----
Qy	829	AsnProValSerLeuProLeuProAspIleAspLeuProAspTPArgHisGlyCysIlePro	848		-----
Db	2282	AACCCAGTGTCACTTCCCAGACTTACCCAGCTGCTGCTGCTGATCTGCTGTTGCTT	2341		-----
Qy	849	CYSGlnAsnMetGluLeuPheAlaLeuCysIleGluThrSerHistYralAlaPhe	868		-----
Db	2342	TGCCAGAAATGAGATTAATGCTGATGAAACAGCCACTATGCTGATGAA	2401		-----
Qy	869	VallysTyrGlyLysAspPheAspSerAlaIlePheAspSerMetAlaAspArgAsp	888		-----
Db	2402	GTGAAGTAGGAGGAGGAGGATCTGCTGCTCTCTGACATGGGATCGGAT	2461		-----
Qy	889	GlyGlyGlnAsnGlyPheAsnIleProGlnValThrProCysProGluValGlyGluYr	908		-----
Db	2462	GCTGGTCAGAAATGGCTCAACATTCCTCAAGTCACCCATGCCAGAGTAGGAGACTAC	2521		-----
Qy	909	LeuLyMetAspSerLeuGluAspIleIleSerLeuAspSerArgGargIleGlyCysAla	928		-----
Db	2522	TGAGATGTTCTGAAAGACTGCATCTGGACTCCAGAGAATCCAGGCTGTGCA	2581		-----
Qy	929	ArgArgLeuLeuCysBspAlaIleYrMetCysMetYrgInserProHmetSerIeYr	948		-----
Db	2582	CGAAGACTGTTGTGATGATATGCTGAGCTTCTGATCTGGACTCCAGAGAATCCAGGCTGTGCA	2641		-----
Qy	949	Lys 949			-----
Db	2642	AAA 2644			-----
RESULT	10				Rat Genome Sequencing Consortium
AC123449					Direct Submission
LOCUS	AC123449	241990 bp	DNA	linear	Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
DEFINITION	Rattus norvegicus clone CH230-21D6, *** SEQUENCING IN PROGRESS *** , 3 unordered pieces.				Submitted (21-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
ACCESSION	AC123449	GT:232654 06			On Sep 21, 2002 this sequence version replaced gi:21903787. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using ATLAS (http://www.hgsc.bcm.edu/projects/rat/). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
VERSION	AC123449_3	HTGS PHASE1	HTGS DRAFT	HTGS ENRICHED	-----
KEYWORDS					-----
SOURCE					-----
ORGANISM	Rattus norvegicus	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			-----
REFERENCE	1 (bases 1 to 241990)				Center: Baylor College of Medicine
AUTHORS	Muzey, D. Marie., Metzker, M. Lee., Abramzon, S., Adams, C., Alder, J., Allen, C., Allen, H., Alstrooks, S., Amin, A., Anguiano, D.,				Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu				
Center project name: GXOF				
Center clone name: CH230-21D6				
----- Summary Statistics -----				
Assembly program: Phrap; version 0.990329				
Consensus quality: 22902 bases at least Q40				
Consensus quality: 231763 bases at least Q30				
Consensus quality: 233057 bases at least Q20				
Estimated insert size: 260790; sum-of-contigs estimation				
Quality coverage: 3x in Q20 bases; sum-of-contigs estimation				
* NOTE: Estimated insert size may differ from sequence length * (see http://www.hgsc.bcm.tmc.edu/docs/Gembank_draft_data.html)				
* NOTE: This sequence may differ more than one clone.				
* NOTE: This is a 'working draft' sequence. It currently				
* consists of 3 contigs. The true order of the pieces				
* is not known and their order in this sequence record is				
* arbitrary. Gaps between the contigs are represented as				
* runs of N, but the exact sizes of the gaps are unknown.				
* This record will be updated with the finished sequence				
* as soon as it is available and the accession number will				
* be preserved.				
* 1 238985: contig of 238985 bp in length				
* 238986 239085: gap of unknown length				
* 239086 240731: contig of 1646 bp in length				
* 240732 240831: gap of unknown length.				
* 240832 241990: contig of 1159 bp in length.				
FEATURES	source	Location/Qualifiers		
	1..241990	/organism="Rattus norvegicus" /mol_type="genomic DNA" /db_xref="taxon:10116" /clone="CH23-0-21D6" 1323..2205 /note="clonal_boundary clone_end_sp6 site_EcoRI end_sequence: BH269859" complement(236297..237124) /note="clone_boundary clone_end:r77 site_EcoRI end_sequence: BH269858"		
misc_feature				
ORIGIN				
Alignment Scores:				
pred. No.:	0	Length:	241990	
Score:	4600.50	Matches:	883	
Percent Similarity:	95.29%	Conservative:	27	
Best Local Similarity:	92.46%	Mismatches:	38	
Query Match:	91.39%	Indels:	8	
DB:	2	Gaps:	5	
US-09-671-687A-3 (1-949) × AC123449 (1-241990)				
Qy	1 MetSerGlyLeuTrpSerGlnGluysValThrSerProTyptpGluGluArgIle 20			
Db	205375 ATGAGTTCGGCTTGCAACCGAGAAAGTTACTTACCTACTGGAAAGACGGGT 205434			
Qy	21 PheTyrIleLeuGlnGluCysSerValThrAspLysGlnThrGlnLysLeuLeuLys 40			
Db	205335 TTTTATCGTTCTCAAGAAATCGAGTAAC---AACAGACACAAGCTCTGAAG 205490			
Qy	61 ProSerAlaLysGlyLeuGlnIleGlyLeuLysIleLeuGlnProHisAla 80			
Db	205551 CCTTCAGTAAAGGCCAAACTATGGAGCATGGCTCACGTCACACTCTCTGTGC 2065690			
Qy	81 ValLeuPheValAspGluIleAsnThrAlaProValGlnGluLeuLeu 99			
Db	205611 GTTCCTGTTGTTGATGAAAGATGTTGAAATTCAAGAGTTACTG 205670			
Qy	100 LeuAlaIleThrAsnCysGluGluArgPheserLeuPheIleAsnArgLeuSer 119			
Db	205671 TTGGCAATTACCAACTCTGAGGAGCCTACGCCATTCCGAAATTCGACTAAGT 205730			
Qy	120 LysGlyLeuGlnIleAspValGlyCysProValysValGlnLeuArgSerGlyGluGlu 139			
Db	205731 AAAGGCTCTCGGAGCAGGCGGACTCTGTGAGAGTAGCTGTCATCTGGAGGAG 205790			
Qy	140 LysPheProGlyValValArgPheArgGlyProLeuLeuAlaGluArgThrValSerGly 159			
Db	205791 AAGTTCCAGAGTGTACGTTCTAGGACTTATTAGCGAGAGCAGGTGTCGG 205850			
Qy	160 IlePhePheGlyValGluIleLeuGluGlyArgGlyGlnGlyPheThrAspGlyVal 179			
Db	205851 ATTTCCTTGAGATGAAATTACTGGAGAAAGCTGCGCCAAAGGTTCACTGATGGCTG 205910			
Qy	180 TyrGlyLysGlnIlePheGlnCysAspGluAspCysGly---PheValAlaLeuAsp 198			
Db	205911 TATCAGGAAAACGCTCTCCAGCTGATGAGGGACTGTGTTGTCATTTGAC 205970			
Qy	199 LysIleGluLeuIleGluAspAspAspPhePheAspIleGluUserAspTyrAlaGlyProGly 218			
Db	205971 AAGCTGGAGCTTATGAAGATGATAACAATGGTTGAAAGTGTATTCAGGCCCCAGGA 206030			
Qy	219 AspThrMetGlnValGluLeuProLeuGluIleAsnSerArgValSerLeuLysGly 238			
Db	206031 GATACAGTCAGGTTGAACCTCCCTCTGAAATTAAACTCCAGAATTCTTGACGTT 206090			
Qy	239 GlyGlyThrIleGluSerGlyThrValIlePheCysAspValLeuProGlyLysGluSer 258			
Db	206091 GGAGGAAGTACAGATTCTGGAAACGTGATATTCTGTGATTTACAGGAAAGAGAGT 206150			
Qy	259 LeuGlyTyrPheValGlyValAspMetAspAsnProIleGlyAsnTrpAspGlyArgPhe 278			
Db	206151 CTAGGATAATTGTTGTTGACATGGATGAACTTACGGAACTTGGATGAACTGTT 206210			
Qy	279 AspGlyVal---IeuCysSerPheAlaCysValGluSerIleLeuHisIleAsn 297			
Db	206211 GATGCCGGTCAGCGGTCAAGCTTTCAGCTCATCTAGTCATTTCCPACACATCAAT 206270			
Qy	298 AspIleLeuProGluSerValThrGlnGluArgArgProProLysLeuAlaPheMetSer 317			
Db	206271 GACATCATCCAGATAGCGTGAACAGAAAGGAACTCCAAACTTGCTTTATGTCA 206330			
Qy	318 ArgGlyValGlyAspLysGlySerSerSerHisAsnLysProLysAlaThrGlySerThr 337			
Db	206331 AGAGCTGGAGTGTGACAAAGCTCATCTAGTCATATAAAACCAACGTTACGGATCTTAC 2063390			
Qy	338 SeraspProGlyAsnArg---ArgSerGluLeuPheTyrThrLeuAsnGlySerSerVal 356			
Db	206391 TCAGGCCCTGGAAAGTAGAAACAGATCTGAATTATTATCCRATTAATGCTTACATCTGTT 206450			
Qy	357 AspSerGlnProGlnSerIleSsSerLysAsnThrTrpTrpIleAspLluvalAlaGluAsp 376			
Db	206451 GACTGACRAACAACTCCAGAAACCCTATGTCAGATGCTCAAGTGGCAAGAC 206510			
Qy	377 ProAlaYsserLeuThrGluIleSerLeuThrGluAsnArgPheHissSerLeuProProSerIle 416			
Db	206571 CCTCCCTCCATGAACTCCCTGTCAGCGAGAACGATTCACCTTACCCCTCCGCCCTG 206630			
Qy	417 ThrlysMetProAsnThrAsnGlySerIleGlyHisSerProLeuSerIeAlaG 436			
Db	206631 ACAGAGATGCCAAACTATGGAGCATGGCTCACGTCACACTCTCTGTGC 206690			
Qy	436 InSerValMetGluGluLeuAsnThrAlaProValGlnGluSerProLeuAlaMetP 456			

Db	206691	AGTCCTGATGGGGAGCTGAACAGCACGCCCTGTCCAGGAGACTCCCCCTGCCAGCT	206750	Qy	456	roProGlyAlaSerHisGlyLeuGluValGlySerIleuAlaGluValIleuAanProp	476	Qy	816	InValHisLeuHisProLyshargLeuAsnHisLysTyrAsnProValSerLeuProLySA	836				
Qy	206751	CTTCCTGGAAATGGACACGGGCTAGGTGGCTCACTGGCTGAAGTAAGAGAACCCC	206810	Db	207831	AGGTTACCTCTATCCAGAGACTAACATCAGATCATCTTACATCTCCAAAG	207890	Db	207831	AGGTTACCTCTATCCAGAGACTAACATCAGATCATCTTACATCTCCAAAG	207890				
Db	206811	CGTTCTATGGGGTTACCGTGGATGGCCAGGCCCTCACTGACCTGCTGCTG	206870	Qy	476	roPheTerGlyValIleArgTrpIleGlyGlnProProGlyLeuAlaLeuAlaG	496	Qy	836	spleuProAspTrpAspTrpArgHisLysGlyCysIleProCysGlnAsnMetGluLeuPheA	856				
Qy	206871	GATTGAACCAGGAATGATAATGCAGGTTCACGACTTGAACTTCAGCTGCTG	206930	Db	207891	ACTTGGCCGACTGGGACTGGACATGGCATCCATGTCAAGATGGACTTATTG	207950	Db	207891	ACTTGGCCGACTGGGACTGGACATGGCATCCATGTCAAGATGGACTTATTG	207950				
Qy	496	IYLeuGluLeuGluAspGluCysAlaGlyCysThrAspGlyThrPheArgLeuArg	516	Qy	856	IvalLeuCysIleGluThrSerHistYryValAlaPheValYstYrgIlySaspapS	876	Qy	856	IvalLeuCysIleGluThrSerHistYryValAlaPheValYstYrgIlySaspapS	876				
Db	206931	ATTCACCTGTGCCCTGAAAGAACGACTGTGAAACTGAGCTGGACCACTG	206990	Db	207951	CTGTGCTCTGATGAAACGCCACATGTTGAGTGGAACTGGGAAAGTGTGACT	208010	Db	207951	CTGTGCTCTGATGAAACGCCACATGTTGAGTGGAACTGGGAAAGTGTGACT	208010				
Qy	536	erArgPheAlaSerLeuGlnProValSerAsnGlnIleGluArgCysAsnSerLeuAlaP	556	Qy	876	erAlaTrpLeuPhePheAspSerMetAlaAspArgAspGlyGlnAsnGlyPheAsnI	896	Qy	876	erAlaTrpLeuPhePheAspSerMetAlaAspArgAspGlyGlnAsnGlyPheAsnI	896				
Db	206991	CTAGGTTGTGATCCTTCAGCTGTTCCANTCAGTCGAAGGGCTAACTCTTAGCAT	207050	Db	208011	CTGCCCTGGCTCTGATGAGCTGGATCTGATGAGCTGGCTAC	208070	Db	208011	CTGCCCTGGCTCTGATGAGCTGGATCTGATGAGCTGGCTAC	208070				
Qy	556	heGlyGlyTyrIleuSerGlyAlaValGluAsnThrProProlysmetGluYsGluG	576	Qy	916	euHISerLeuAspSerArgArgIleGlnGlyCysAlaArgGluLeuCysAspAlaT	936	Qy	916	euHISerLeuAspSerArgArgIleGlnGlyCysAlaArgGluLeuCysAspAlaT	936				
Db	207051	TTGGGCTCACTTAAGTAAGTAGTGAAGAAAATGCACTAAATGAAAGAGA	207110	Db	208131	TGCACCTTTGACTCCAGGATCAAGGTCAACGAGGATCAAGGTCAACGAGT	208190	Db	208131	TGCACCTTTGACTCCAGGATCAAGGTCAACGAGT	208190				
Qy	576	IYLeuGluIleMetIleGlyIlySlysIleGlyIleGlnGlyIleGlnSerCystYI	596	Qy	936	YrmCysSmtryGlnSerProThrMetSerLeutYrLys	949	Qy	936	YrmCysSmtryGlnSerProThrMetSerLeutYrLys	949				
Db	207111	GTTTAGATGATGTTGAAAGGATCAGGGCCATTACAATCTGTGTA	207170	Db	208191	ACATGGCATGACAGTCACATGAGTTGACAA	208231	Db	208191	ACATGGCATGACAGTCACATGAGTTGACAA	208231				
RESULT 11															
AC098162/c															
LOCUS		AC098162		251132 bp		DNA		linear							
DEFINITION		CH230-3014, *** SEQUENCING IN PROGRESS ***,													
Rattus norvegicus clone CH230-3014, *** SEQUENCING IN PROGRESS ***,															
3 unordered pieces.		AC098162		251132 bp		CH230-3014, *** SEQUENCING IN PROGRESS ***,									
AC098162.7															
VERSION		GI:30522215		GI:30522215		HTG; HTGS_PHASB1; HTGS_DRAFT; HTGS_ENRICHED.									
Rattus norvegicus (Norway rat)															
KEYWORDS		Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;		Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;		Mammalia; Eutheria; Rodentia; Muridae; Murinae;									
Rattus norvegicus															
ORGANISM		1 (bases 1 to 251132)		1 (bases 1 to 251132)		Rattus norvegicus									
Rattus norvegicus															
REFERENCE		AC098162		AC098162		AC098162									
AUTHORS															
Allan, C., Allen, H., Alabrooks, S., Amin, A., Anguiano, D., Anyelbechi, V., Ayoggi, A., Ayodeji, M., Baca, B., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Biswalo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Bunay, C., Burch, P., Burrell, K., Calderon, B., Cardenas, V., Carter, K., Cavaros, I., Cesair, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Cleveland, C., Cockrell, R., Cox, C., Coyne, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Andra, C., Dederich, D., Delgado, O., Denson, S., Dernano, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dusun-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escott, M., Eugene, C., Evans, C.A., Failes, T., Fan, G., Fernandez, S., Finley, S., Flagg, N., Forbes, L., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Gara, M., Gebregeorgis, E., Gear, K., Gill, R., Kelly, S., King, L., Kovar, C., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hernandez, J.J., Harvey, Y., Havlak, P., Hawes, A., Henderison, N., Hernandez, R., Hines, S., Hladun, S.L., Hodges, M., Hollings, B., Howell, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, P., Jolivet, A., Karpashy, S., Kelly, S., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Louissegard, H., Lozada, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindarree, M., Maloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.R.Z., Meenan, E., Milosevijevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidas, M., Murphy, M., Nair, L.,															
AC098162															
ACCESSION		AC098162.7		GI:30522215		HTG; HTGS_PHASB1; HTGS_DRAFT; HTGS_ENRICHED.									
Rattus norvegicus (Norway rat)															
VERSION		Rattus norvegicus		Rattus norvegicus		Rattus norvegicus									
Rattus norvegicus															
KEYWORDS		Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;		Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;		Mammalia; Eutheria; Rodentia; Muridae; Murinae;									
Rattus norvegicus															
SOURCE		Rattus norvegicus		Rattus norvegicus		Rattus norvegicus									
Rattus norvegicus															
ORGANISM		Rattus norvegicus		Rattus norvegicus		Rattus norvegicus									
Rattus norvegicus															
REFERENCE		Rattus norvegicus		Rattus norvegicus		Rattus norvegicus									
Rattus norvegicus															
AUTHORS		Rattus norvegicus		Rattus norvegicus		Rattus norvegicus									
Rattus norvegicus															
VERSION		Rattus norvegicus		Rattus norvegicus		Rattus norvegicus									
Rattus norvegicus															
KEYWORDS		Rattus norvegicus		Rattus norvegicus		Rattus norvegicus									
Rattus norvegicus															
SOURCE		Rattus norvegicus		Rattus norvegicus		Rattus norvegicus									
Rattus norvegicus															
ORGANISM		Rattus norvegicus		Rattus norvegicus		Rattus norvegicus									
Rattus norvegicus															
REFERENCE		Rattus norvegicus		Rattus norvegicus		Rattus norvegicus									
Rattus norvegicus															
AUTHORS		Rattus norvegicus		Rattus norvegicus		Rattus norvegicus									
Rattus norvegicus															
VERSION		Rattus norvegicus		Rattus norvegicus		Rattus norvegicus									
Rattus norvegicus															
KEYWORDS		Rattus norvegicus		Rattus norvegicus		Rattus norvegicus									
Rattus norvegicus															
SOURCE		Rattus norvegicus		Rattus norvegicus		Rattus norvegicus									
Rattus norvegicus															
ORGANISM		Rattus norvegicus		Rattus norvegicus		Rattus norvegicus									
Rattus norvegicus															
REFERENCE		Rattus norvegicus		Rattus norvegicus		Rattus norvegicus									
Rattus norvegicus															
VERSION		Rattus norvegicus		Rattus norvegicus		Rattus norvegicus									
Rattus norvegicus															
KEYWORDS		Rattus norvegicus		Rattus norvegicus		Rattus norvegicus									
Rattus norvegicus															
SOURCE		Rattus norvegicus		Rattus norvegicus		Rattus norvegicus									
Rattus norvegicus															
ORGANISM		Rattus norvegicus		Rattus norvegicus		Rattus norvegicus									
Rattus norvegicus															
REFERENCE		Rattus norvegicus		Rattus norvegicus		Rattus norvegicus									
Rattus norvegicus															
VERSION		Rattus norvegicus		Rattus norvegicus		Rattus norvegicus									
Rattus norvegicus															
KEYWORDS		Rattus norvegicus		Rattus norvegicus											

Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,	*	248387	contig of 248387 bp in length
Okwonou,G., Olarmpungsagoon,A., Pal,S., Parks,K.,	*	248388	gap of unknown length
Pasternak,S., Paul,H., Perez,L., Pfannkoch,C.,	*	248388	contig of 1049 bp in length
Poizner,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,	*	249337	gap of unknown length
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Reuter,M.A., Reigh,R.,	*	249337	contig of 1496 bp in length.
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,P.,			LocationQualifiers
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,			1..251132
Sanders,W., Savery,G., Scherer,S., Scott,G., Shataman,S., Shen,H.,			/organism="Rattus norvegicus"
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,			/mol type="genomic DNA"
Sneed,A., Soderren,E., Song,X.-Z., Sorelle,R., Soza,J.,			/db_Xref="taxon:10116"
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,			/clone="CH230-3014"
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usman,R.,			1..1846
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.-J.,			/note="ngs_end_extension
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,			clone_end:Sp6"
Williams,G., Wilson,R., Wleczek,R., Woeden,H., Worley,K.,			/note="ngs_end_extension
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,			clone_end:Sp6"
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von			complement(6517..7123)
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,			/note="clone_boundary
Weinstock,G. and Gibbs,R.A.			clone_end:sp6
Direct Submission			site:EcrRI
Unpublished			end_sequence:BH2925593"
			ORIGIN
			Alignment Scores:
			Pred. No.:
			Score:
		0	Length:
		4600..50	Matches:
		95..29%	Conservative:
		92..46%	Percent Similarity:
		91..39%	Best Local Similarity:
		2	Query Match:
		5	DB:
			US-09-671-687A-3 (1..949) x AC098162 (1..251132)
			COMMENT
			The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by size gaps filled with Ns (to the estimated size). The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
			----- Genome Center
			Center: Baylor College of Medicine
			Center code: BCM
			Web site: http://www.hgscbcm.tmc.edu/
			Contact: hgsc-help@bcm.tmc.edu
			----- Project Information
			Center project name: GGWL
			Center clone name: CH230-3014
			----- Summary Statistics
			Assembly program: Atlas 3.0;
			Consensus quality: 241087 bases at least Q40
			Consensus quality: 243305 bases at least Q30
			Consensus quality: 244914 bases at least Q20
			Estimated insert size: 256850; sum-of-contigs estimation
			Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
			* NOTE: Estimated insert size may differ from sequence length *(see http://www.hgscbcm.tmc.edu/docs/Genbank_draft_data.html)
			* NOTE: This sequence may represent more than one clone.
			* NOTE: This is a 'working draft' sequence. It currently
			* consists of 3 contigs. The true order of the pieces
			* is not known and their order in this sequence record is
			* arbitrary. Gaps between the contigs are represented as
			* runs of N, but the exact sizes of the gaps are unknown.
			* This record will be updated with the finished sequence
			* as soon as it is available and the accession number will
			* be preserved.

		misc_feature	1. .2341 /note="highly similar to AB020656 Homo sapiens mRNA for KIAA0849 protein"
Db	1665	AGAGAATGCTACGACGATCGGACATCTCACTGGAAAAATAAAGCAATTGTAAACC	1724
Qy	813	CysAsnThrGlnValHisLeuHisProLysArgLeuAsnSlysTyrAsnProValSer	832
Db	1725	TGCAACACTCAAGTCACCTTCATCCGAAGCCTGATCATAAATACCAGTGTC	1784
Qy	833	LeuProLysAspLeuProAspPtpAspTrpArgHisGlyCysIleProCysGlnAsnMet	852
Db	1785	CTTCCTCAAAGACTTACCGGACTGGGACTGGGACTGGGACTGGGACTGGGACTGGG	1844
Qy	853	GluLeuPheAlaValLeuCystileGluThrSerHistYrrValAlaPheValYstYrgY	872
Db	1845	GAGTTATTGCTGTTCTGCTGATAGAAACAAGCCACTATGTCAGTATGGG	1904
Qy	873	LysAspAspSerAlaAlaPheAspSerMetAlaAlaPheAspSerMetAlaAlaPheAsp	892
Db	1905	AAGGACGATTCGCTCTGGCTCTGGATCTGGATCTGGATCTGGATCTGGATCTGGAA	1964
Qy	893	GlyPheAsnIleProGlnValThrProCysProGluValGlyGluGlyLeuIleMetSer	912
Db	1965	GGCTTCACATCCTCAAGTACCCATGCCGATGCCGATGCCGATGCCGATGCCGATGCC	2024
Qy	913	LeuGluAspLeuAspSerArgIleGlnGlyCysAlaArgAlaGlnGlyCAGACTGAGATGTC	932
Db	2025	CTGGAAAGACCTGCATTCCTGACTCCAGGAAATCCAAAGCTGTGACAGAACAGACTGCT	2084
Qy	933	CysAspAlaTyrMetCysMetTyrGlnSerProThrMetSerLeuTyrLys	949
Db	2085	TGTGATSCATATATGTCATGTACAGAGTCCAAACATGAGTTGTACAAA	2135
RESULT 13			
LOCUS	AK000187	AK000187	2341 bp mRNA linear PRI 13-SEP-2003
DEFINITION		Homo sapiens cDNA FLJ20180 f1s	, clone COL1238, highly similar to KIAA0849 protein.
ACCESSION	AB020656	Homo sapiens mRNA for KIAA0849 protein.	
VERSION	AK000187	GI:7020107	
KEYWORDS	AK000187.1	GI:7020107	
ORGANISM		Oligo capping; f1s (full insert sequence).	
SOURCE	Homo sapiens (human)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikeda,Y., Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.		
AUTHORS		Unpublished (bases 1 to 2341)	
JOURNAL		Sugano,S., Suzuki,Y., Ora,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.	
REFERENCE		Shibahara,T., Tanaka,T. and Nakamura,Y.	
AUTHORS		Direct Submission	
JOURNAL		Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Department of Virology; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail: fcdna@ims.u-tokyo.ac.jp, Tel: 81-3-5449-5286, Fax: 81-3-5449-5416)	
COMMENT		NEPO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology cDNA library construction, 5' - & 3' -end one pass sequencing; Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).	
FEATURES	source	Location/Qualifiers	
		1. .2341	
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref=taxon:9606	
		/clone="COL10238"	
		/tissue_type="colon"	
		/clone.Lib="COL"	
		/note="cloning vector pME18SFL3"	

